

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2003, 11:34:34 ; Search time 732 Seconds
(without alignments)
9546.674 Million cell updates/sec

Title: US-09-782-745-14
Perfect score: 538
Sequence: 1 ACGCCAGGAGCTGTGAGGC.....CTGCCAAGAAAAA 538

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1994485 seqs, 6494577260 residues
Total number of hits satisfying chosen parameters: 3988970

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl NoHTG:*

- 1: gb_da:*
- 2: gb_in:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_pat:*
- 6: gb_ph:*
- 7: gb_pl:*
- 8: gb_pr:*
- 9: gb_ro:*
- 10: gb_sts:*
- 11: gb_sy:*
- 12: gb_un:*
- 13: gb_vt:*
- 14: em_ba:*
- 15: em_fun:*
- 16: em_hum:*
- 17: em_in:*
- 18: em_mu:*
- 19: em_om:*
- 20: em_or:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_un:*
- 28: em_vi:*
- 29: em_htg_hum:*
- 30: em_htg_inv:*
- 31: em_htg_other:*
- 32: em_htg_mus:*
- 33: em_htg_pin:*
- 34: em_htg_rod:*
- 35: em_htg_mam:*
- 36: em_htg_vrt:*
- 37: em_sy:*
- 38: em_htgo_hum:*
- 39: em_htgo_mus:*
- 40: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	538	100.0	538	5	AR028488	AR028488 Sequence
2	538	100.0	538	5	I55851	I55851 Sequence 14
3	533.2	99.1	551	8	BC018052	BC018052 Homo sapi
4	531.6	98.8	1245	5	AX285022	AX285022 Sequence
5	530	98.5	530	8	HSU19143	U19143 Human GAGE-
6	525.4	97.7	528	8	AF055473	AF055473 Homo sapi
7	512.8	95.3	540	5	AR028490	AR028490 Sequence
8	512.8	95.3	540	5	I55853	I55853 Sequence 16
9	512.2	95.2	637	8	BC024914	BC024914 Homo sapi
10	510.2	94.8	539	5	AR028492	AR028492 Sequence
11	510.2	94.8	539	5	I55855	I55855 Sequence 18
12	506	94.1	606	8	BC031628	BC031628 Homo sapi
13	504.8	93.8	532	5	AR028491	AR028491 Sequence
14	504.8	93.8	532	5	I55854	I55854 Sequence 17
15	501.4	93.2	528	5	AX334151	AX334151 Sequence
16	501.4	93.2	528	8	HSU19145	U19145 Human GAGE-
17	498.8	92.7	527	8	HSU19147	U19147 Human GAGE-
18	497.6	92.5	524	8	AF058988	AF058988 Homo sapi
19	496.8	92.3	524	8	HSU19146	U19146 Human GAGE-
20	488	90.7	526	8	AF055474	AF055474 Homo sapi
21	429.2	79.8	560	5	AR028489	AR028489 Sequence
22	429.2	79.8	560	5	I55852	I55852 Sequence 15
23	421.2	78.3	552	8	HSU19144	U19144 Human GAGE-
24	374.4	69.6	646	5	AR028482	AR028482 Sequence
25	374.4	69.6	646	5	I55845	I55845 Sequence 1
26	374.4	69.6	646	8	HSU19142	U19142 Human GAGE-
27	374.4	69.6	648	5	I57317	I57317 Sequence 1
28	301.4	56.0	530	5	AX284300	AX284300 Sequence
29	299.2	55.6	365	5	AX284692	AX284692 Sequence
30	209.6	39.0	662	8	BC004861	BC004861 Homo sapi
31	209.6	39.0	676	8	AF058989	AF058989 Homo sapi
32	185.8	34.5	648	8	BC009232	BC009232 Homo sapi
33	183.8	34.2	642	5	AX359705	AX359705 Sequence
34	182.4	33.9	493	8	HSA318881	AJ318881 Homo sapi
35	179.8	33.4	620	8	HSA318880	AJ318880 Homo sapi
36	173.8	32.3	611	5	AX226501	AX226501 Sequence
37	163.2	30.3	79539	8	AC093664	AC093664 Homo sapi
38	156	29.0	580	5	AX078298	AX078298 Sequence
39	146.8	27.3	763	8	BC009230	BC009230 Homo sapi
40	140.8	26.2	20587	8	HSL185E6A	Z68274 Human DNA s
41	140.8	26.2	62493	8	HSL193G15	AL117391 Human DNA
42	140.8	26.2	62493	8	HSL193G15	AL096838 Human DNA
43	131.4	24.4	9531	8	AF055475	AF055475 Homo sapi
44	130.4	24.2	475	5	AX226497	AX226497 Sequence
45	129.8	24.1	194418	8	AF235098	AF235098 Homo sapi

ALIGNMENTS

RESULT 1
AR028488
LOCUS AR028488
DEFINITION Sequence 14 from patent US 5858689.
ACCESSION AR028488
VERSION AR028488.1 GI:5940461
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 538)
AUTHORS van der Bruggen, P., van den Eynde, B., DeBacker, O. and Boon-Falleur, T.
TITLE Isolated peptides derived from the gage tumor rejection antigen precursor and uses thereof
JOURNAL Patent: US 5858689-A 14 12-JAN-1999;

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAX Plate: 32 Row: k Column: 1
This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 4503878.

FEATURES
source
1. .551
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="LocusID:26749"
/db_xref="taxon:9606"
/clone="MGC:26395 IMAGE:4812462"
/tissue_type="Brain, hippocampus"
/clone_lib="NIH MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
97.447
/codon_start=1
/product="G antigen 8"
/protein_id="AAH18052.1"
/db_xref="GI:17390106"
/translation="MSMRGRSTVPRPRYVERPEMIGPMRPFQSFDEVFATPERGE
PATQRQDPAAAGDEBGASAGGPRPEADSQEGHPQTGCECEDGPDGQEMDPNPE
EVKPEEGEKQSQ"
BASE COUNT 164 a 118 c 160 g 109 t
ORIGIN

Query Match 99.1%; Score 533.2; DB 8; Length 551;
Best Local Similarity 99.4%; Pred. No. 1.3e-122;
Matches 535; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AGCCAGGAGCTGTGAGGAGCTGTGTGGTTCCTCGCGTCCGGACTCTTTTTCCTCT 60
Db 14 AGCCAGGAGCTGTGAGGAGCTGTGTGGTTCCTCGCGTCCGGACTCTTTTTCCTCT 73
QY 61 ACTGAGATTCATCTGTGTAATATGAGTTGGCGAGGAATCGACCTATCGGCTAGAC 120
Db 74 ACTGAGATTCATCTGTGTAATATGAGTTGGCGAGGAATCGACCTATCGGCTAGAC 133
QY 121 CAAGACGCTACGTAGAGCTCTGTAATATGAGTTGGCGCTATGCGGCTAGAGT 180
Db 134 CAAGACGCTACGTAGAGCTCTGTAATATGAGTTGGCGCTATGCGGCTAGAGT 193
QY 181 ATGAAGTGAACCCAGCAACACCTGAAAGAGGGAACAGCAACTCAACGTCAGGATCCTG 240
Db 194 ATGAAGTGAACCCAGCAACACCTGAAAGAGGGAACAGCAACTCAACGTCAGGATCCTG 253
QY 241 CAGCTGCTCAGGAGGAGGATGAGGAGCATCTCGAGTCAAGGCGGAGCCTGAAG 300
Db 254 CAGCTGCTCAGGAGGAGGATGAGGAGCATCTCGAGTCAAGGCGGAGCCTGAAG 313
QY 301 CTCATAGCCAGGAACAGGGTCAACCAAGAGTGGGTGAGTGAAGATGCTCTGATG 360
Db 314 CTGATAGCCAGGAACAGGGTCAACCAAGAGTGGGTGAGTGAAGATGCTCTGATG 373
QY 361 GGCAGGAGATGGAACCGCAATCCAGAGAGGTGAAACCGCTGAAAGAGGTGAAAGC 420
Db 374 GGCAGGAGATGGAACCGCAATCCAGAGAGGTGAAACCGCTGAAAGAGGTGAAAGC 433
QY 421 AATCAGCTGTAAAGAGACAGTTGAAATGATGATGATGATGATGATGATGATGAT 480
Db 434 AATCAGCTGTAAAGAGACAGTTGAAATGATGATGATGATGATGATGATGATGAT 493
QY 481 TGTTCAATTAATTTCTCCCAATAAAGCTTTACAGCTTCTGCAAGAAAAA 538
Db 494 TGTTCAATTAATTTCTCCCAATAAAGCTTTACAGCTTCTGCAAGAAAAA 551

RESULT 4
AX285022
LOCUS
DEFINITION
Sequence 827 from Patent W00179556.
ACCESSION
AX285022
VERSION
AX285022.1 GI:17045710
KEYWORDS
human.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
Lillie, J., Brown, J. L., Bolt, A. and van Huffel, C.
Novel genes, compositions and methods for the identification,
assessment, prevention, and therapy of human cancers
Patent: WO 0179556-A 827 25-OCT-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES
source
1. .1245
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 382 a 308 c 390 g 153 t 12 others
ORIGIN

Query Match 98.8%; Score 531.6; DB 5; Length 1245;
Best Local Similarity 99.3%; Pred. No. 3.4e-122;
Matches 534; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACGCCAGGAGCTGTGAGGAGCTGTGTGGTTCCTCGCGTCCGGACTCTTTTTCCTCT 60
Db 179 ACGCCAGGAGCTGTGAGGAGCTGTGTGGTTCCTCGCGTCCGGACTCTTTTTCCTCT 238
QY 61 ACTGAGATTCATCTGTGTAATATGAGTTGGCGAGGAATCGACCTATCGGCTAGAC 120
Db 239 ACTGAGATTCATCTGTGTAATATGAGTTGGCGAGGAATCGACCTATCGGCTAGAC 298
QY 121 CAAGACGCTACGTAGAGCTCTGTAATATGAGTTGGCGCTATGCGGCTAGAGT 180
Db 299 CAAGACGCTACGTAGAGCTCTGTAATATGAGTTGGCGCTATGCGGCTAGAGT 358
QY 181 ATGAAGTGAACCCAGCAACACCTGAAAGAGGGAACAGCAACTCAACGTCAGGATCCTG 240
Db 359 ATGAAGTGAACCCAGCAACACCTGAAAGAGGGAACAGCAACTCAACGTCAGGATCCTG 418
QY 241 CAGCTGCTCAGGAGGAGGATGAGGAGCATCTGAGGTCAGGTCAGGCGGAGCCTGAAG 300
Db 419 CAGCTGCTCAGGAGGAGGATGAGGAGCATCTGAGGTCAGGTCAGGCGGAGCCTGAAG 478
QY 301 CTCATAGCCAGGAACAGGGTCAACCAAGAGTGGGTGAGTGAAGATGCTCTGATG 360
Db 479 CTCATAGCCAGGAACAGGGTCAACCAAGAGTGGGTGAGTGAAGATGCTCTGATG 538
QY 361 GGCAGGAGATGGAACCGCAATCCAGAGAGGTGAAACCGCTGAAAGAGGTGAAAGC 420
Db 539 GGCAGGAGATGGAACCGCAATCCAGAGAGGTGAAACCGCTGAAAGAGGTGAAAGC 598
QY 421 AATCAGCTGTAAAGAGACAGTTGAAATGATGATGATGATGATGATGATGATGAT 480
Db 599 AATCAGCTGTAAAGAGACAGTTGAAATGATGATGATGATGATGATGATGATGAT 658
QY 481 TGTTCAATTAATTTCTCCCAATAAAGCTTTACAGCTTCTGCAAGAAAAA 538
Db 659 TGTTCAATTAATTTCTCCCAATAAAGCTTTACAGCTTCTGCAAGAAAAA 716

RESULT 5
HSU19143
LOCUS
DEFINITION
Human GAGE-2 protein mRNA, complete cds.
ACCESSION
U19143
VERSION
U19143.1 GI:914900
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
A new family of genes coding for an antigen recognized by
autologous cytolytic T lymphocytes on a human melanoma
J. Exp. Med. 182 (3), 689-698 (1995)
JOURNAL
MEDLINE
95378788

PUBMED 7544395
REFERENCE 2 (bases 1 to 530)
AUTHORS Van Den Eynde,B.J.
TITLE Direct Submission
JOURNAL Submitted (28-DEC-1994) Benoit J Van Den Eynde, Ludwig Institute
For Cancer Research, 74 Avenue Hippocrate, BRUSSELS, 1200, BELGIUM
FEATURES Location/Qualifiers
source
1..530
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="female"
/cell_line="MZ2-MEL.43"
/tissue_type="melanoma"
/dev_stage="adult"
84..434
/codon_start=1
/product="GAGE-2 protein"
/protein_id="AA82745.1"
/db_xref="GI:914901"
PATQRDPAAARQEGDEGASAGQPKPEAHSQEQGHQPTGCECEDPGDQEMDPNPE
EVTPEGEKQSQ"
BASE COUNT 152 a 116 c 155 g 107 t
ORIGIN
Query Match 98.5%; Score 530; DB 8; Length 530;
Best Local Similarity 100.0%; Pred. No. 8e-122;
Matches 530; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACGCCAGGAGCTGTGAGCGAGTGTGTGTGTTCTCTGCGTCCGACCTTTTTCCTCT 60
Db 1 ACGCCAGGAGCTGTGAGCGAGTGTGTGTGTTCTCTGCGTCCGACCTTTTTCCTCT 60
QY 61 ACTGAGATTTCATCTGTGTAATATAGTTTGGCGAGGAGATCGACTATCGGCTAGAC 120
Db 61 ACTGAGATTTCATCTGTGTAATATAGTTTGGCGAGGAGATCGACTATCGGCTAGAC 120
QY 121 CAAGACCTAGTAGAGCTCTGAAATGATTGGGCTATGCGGCCGAGCAGTTCAGTG 180
Db 121 CAAGACCTAGTAGAGCTCTGAAATGATTGGGCTATGCGGCCGAGCAGTTCAGTG 180
QY 181 ATGAAGTGGAAACGACAACTGTAAGAAAGGGAAACAGCAACTCAACGTCAGATCCTG 240
Db 181 ATGAAGTGGAAACGACAACTGTAAGAAAGGGAAACAGCAACTCAACGTCAGATCCTG 240
QY 241 CAGTGTCTCAGGAGGAGATGAGGAGCATCTCAGGTCAAGGCCGAGCCTGAAG 300
Db 241 CAGTGTCTCAGGAGGAGATGAGGAGCATCTCAGGTCAAGGCCGAGCCTGAAG 300
QY 301 CTCATAGCCAGGAAACAGGCTCAACACAGACTGGGTGTGAGTGTGAAGATGGTCTGATG 360
Db 301 CTCATAGCCAGGAAACAGGCTCAACACAGACTGGGTGTGAGTGTGAAGATGGTCTGATG 360
QY 361 GCGAGGAGATGGACCCGCCAAATCCAGAGAGGTGAAAACGCTGAAGAGGTGAAAAGC 420
Db 361 GCGAGGAGATGGACCCGCCAAATCCAGAGAGGTGAAAACGCTGAAGAGGTGAAAAGC 420
QY 421 AATCACAGTGTAAAGAGACACGTTGAAATGATGACGCTGCTCTATGTTGGAATTT 480
Db 421 AATCACAGTGTAAAGAGACACGTTGAAATGATGACGCTGCTCTATGTTGGAATTT 480
QY 481 TGTTTCATTAATAATCTCCCAATAAAGCTTTTACAGCCTTTCTGCAAGAAAA 530
Db 481 TGTTTCATTAATAATCTCCCAATAAAGCTTTTACAGCCTTTCTGCAAGAAAA 530
RESULT 6
AF055473
LOCUS Homo sapiens GAGE-8 mRNA, complete cds.
DEFINITION AF055473
ACCESSION AF055473
VERSION AF055473.1 GI:3511022
KEYWORDS

SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 528)
AUTHORS De Backer,O., Arden,K.C., Boretti,M., Vantomme,V., De Smet,C.,
Czekay,S., Viars,C.S., De Plaen,B., Brasseur,F., Chomez,P., Van den
Eynde,B., Boon,T. and van der Bruggen,P.
TITLE Characterization of the GAGE genes that are expressed in various
human cancers and in normal testis
JOURNAL Cancer Res. 59 (13), 3157-3165 (1999)
MEDLINE 99323388
PUBMED 10397259
REFERENCE 2 (bases 1 to 528)
AUTHORS De Backer,O.R.Y.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1998) Ludwig Institute for Cancer Research,
Brussels Branch, 74 av. Hippocrate, Brussels B-1200, Belgium
FEATURES Location/Qualifiers
source
1..528
/organism="Homo sapiens"
/db_xref="taxon:9606"
73..423
/codon_start=1
/product="GAGE-8"
/protein_id="AAC33676.1"
/db_xref="GI:3511023"
/translation="MSWRGRSTYRPRRYVEPPMIGRPEQFSDEVEPATPEEGE
PATQRDPAAARQEGDEGASAGQPKPEADSQEQGHQPTGCECEDPGDQEMDPNPE
EVTPEGEKQSQ"
BASE COUNT 158 a 112 c 151 g 107 t
ORIGIN
Query Match 97.7%; Score 525.4; DB 8; Length 528;
Best Local Similarity 99.8%; Pred. No. 1.1e-120;
Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 12 CTGTGAGGCACTGTGTGTGTTCTCTGCGTCCGACCTTTTTCCTTACTGAGATCA 71
Db 1 CTGTGAGGCACTGTGTGTGTTCTCTGCGTCCGACCTTTTTCCTTACTGAGATCA 60
QY 72 TCTGTGTAATATAGTTTGGCGAGGAGATCGACTATCGGCTAGACCAAGACGCTAC 131
Db 61 TCTGTGTAATATAGTTTGGCGAGGAGATCGACTATCGGCTAGACCAAGACGCTAC 120
QY 132 GTAGAGCTCTGAAATGATTGGGCTATGCGGCCGAGCAGTTCAGTGAAGTGA 191
Db 121 GTAGAGCTCTGAAATGATTGGGCTATGCGGCCGAGCAGTTCAGTGAAGTGA 180
QY 192 CCAGCAACCTGAAAGAGGGAAACAGCAACTCAACGTCAGGATCCTGCACTGCTCAG 251
Db 181 CCAGCAACCTGAAAGAGGGAAACAGCAACTCAACGTCAGGATCCTGCACTGCTCAG 240
QY 252 GAGGAGAGATGAGGAGGAGCATCTGAGGTCAAGGCCGAGGCTGAAGCTCATAGCCAG 311
Db 241 GAGGAGAGATGAGGAGGAGCATCTGAGGTCAAGGCCGAGGCTGAAGCTCATAGCCAG 300
QY 312 GAACAGGCTCAACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTCTGATGGCAGAGATG 371
Db 301 GAACAGGCTCAACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTCTGATGGCAGAGATG 360
QY 372 GACCGCCAAATCCAGAGGAGGTGAAAACGCTTGAAGAGGTGAAAAGCAATCACAGTGT 431
Db 361 GACCGCCAAATCCAGAGGAGGTGAAAACGCTTGAAGAGGTGAAAAGCAATCACAGTGT 420
QY 432 TAAAGAGACACAGCTTGAATGATGACGCTGCTCTATGTTGGAATTTGTTTCATTA 491
Db 421 TAAAGAGACACAGCTTGAATGATGACGCTGCTCTATGTTGGAATTTGTTTCATTA 480
QY 492 ATTCTCCCAATAAAGCTTTTACAGCCTTTCTGCAAGAAAAA 538
Db 481 ATTCTCCCAATAAAGCTTTTACAGCCTTTCTGCAAGAAAAA 527

COMMENT

Contact: MCC help desk
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: ARCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (fLML)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: angbcm.tmc.edu
Gunarathne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

AR028492	AR028492	Sequence 18 from patent US 585689.	539 bp	DNA	linear	PAT 29-SEP-1999
LOCUS						
DEFINITION						
ACCESSION	AR028492					
VERSION	AR028492.1	GI:5940465				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	Unclassified.					
AUTHORS	1 (bases 1 to 539)					
	van der Bruggen, P., van den Eynde, B., DeBacker, O. and					
	Boon-Failleur, T.					
TITLE	Isolated peptides derived from the gage tumor rejection antigen					
	precursor and uses thereof					
JOURNAL	Patent: US 585689-A	18 12-JAN-1999;				
FEATURES	Location/Qualifiers					
source	1..539					
	/organism="unknown"					
BASE COUNT	158 a	113 c	157 g	111 t		
ORIGIN						

RESULT 11
I55855
LOCUS
DEFINITION
Sequence 18 from patent US 5648226.
ACCESSION
I55855
VERSION
I55855.1 GI:2476649
KEYWORDS
Unknown.
SOURCE
Unknown.
ORGANISM
Unclassified.
REFERENCE
1 (bases 1 to 539)
AUTHORS
Van den Eynde, B., DeBacker, O. and Boon-Palleux, T.
TITLE
Isolated peptides derived from tumor rejection antigens, and their use
JOURNAL
Patent: US 5648226-A 18 15-JUL-1997;
FEATURES
Location/Qualifiers
1..539
/organism="unknown"
BASE COUNT 158 a 113 c 157 g 111 t
ORIGIN
Query Match 94.8%; Score 510.2; DB 5; Length 539;
Best Local Similarity 98.0%; Pred. No. 6.9e-117;
Matches 528; Conservative 0; Mismatches 8; Indels 3; Gaps 1;
QY 3 GCCAGGAGCTGTGAGGAGCTGCTGTGTTCTCGCGTCCGGACTCTTTTCTCTCTAC 62
Db 1 GCCAGGAGCTGTGAGGAGCTGCTGTGTTCTCGCGTCCGGACTCTTTTCTCTCTAC 60
QY 63 TGAGATTCATCTGTGTAATATGATGTTGGCAGGAAGATCGACC---TATCGGCTAGA 119
Db 61 TGAGATTCATCTGTGTAATATGATGTTGGCAGGAAGATCGACCTATTATTGGCTAGA 120
QY 120 CCAAGAGCTACGTAGAGCTCTCTGAATGATTTGGGCTATCGGCGCCGAGCAGTTCAGT 179
Db 121 CCAAGGCGCTATGTACAGCTCTCTGAAGTGTATTTGGGCTATCGGCGCCGAGCAGTTCAGT 180
QY 180 GATCAAGTGGAAACAGCAACACCTGAAGAGGGGAAACAGCAACTCAAGCTCAGGATCCT 239
Db 181 GATCAAGTGGAAACAGCAACACCTGAAGAGGGGAAACAGCAACTCAAGCTCAGGATCCT 240
QY 240 GCAGCTGCTCAGGAGGAGAGATGAGGAGCATCTGCAGGTCAAGGGCCGAGCCTGAA 299
Db 241 GCAGCTGCTCAGGAGGAGAGATGAGGAGCATCTGCAGGTCAAGGGCCGAGCCTGAA 300
QY 300 GCTCATGCCAGGACAGGGTCACCCAGACTGGGTGTGAGTGTGAAGATGTCCTGAT 359
Db 301 GCTCATGCCAGGACAGGGTCACCCAGACTGGGTGTGAGTGTGAAGATGTCCTGAT 360
QY 360 GGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAACGCCCTGAAGAAGTGAAAG 419
Db 361 GGGCAGGAGGTGGACCCGCCAAATCCAGAGGAGGTGAAACGCCCTGAAGAAGTGAAAG 420
QY 420 CAATCAGCTGTAAAGAAAGACACGTTGAATGATGCGGCTCTCTATGTTGAAAT 479
Db 421 CAATCAGCTGTAAAGAAAGACACGTTGAATGATGCGGCTCTCTATGTTGAAAT 480
QY 480 TTGTTTCAATAAATCTCCCAATAAGCTTTACAGCTTTCTGCAAGAAAGAAAAA 538
Db 481 TTGTTTCAATAAATCTCCCAATAAGCTTTTACAGCTTTCTGCAAGAAAGAAAAA 539
RESULT 12
BC031628
LOCUS
DEFINITION
Homo sapiens, G antigen 7B, clone MGC:34597 IMAGE:516892, mRNA, complete cds.
ACCESSION
BC031628
VERSION
BC031628.1 GI:21619469
KEYWORDS
MGC.
SOURCE
human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 606)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 51 Row: b Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4503876.
FEATURES
Location/Qualifiers
1..606
/organism="Homo sapiens"
/db_xref="LocustID:26748"
/clone="MGC:34597 IMAGE:516892"
/tissue type="Brain, adult medulla"
/clone lib="NIH MGC_119"
/lab host="DH10B"
/note="Vector: pCMV-SPORT6"
73..426
/codon_start=1
/product="G antigen 7B"
/protein_id="AAH31628.1"
/db_xref="GI:21619470"
/translation="MSWRGRSTYYWPRPRXYVOPMIGMPRQFSDVEPATPEEG
EPATQRDPAAAGEGDEGASAGGPKPEAHSQGHPTGCECEDGPDGEMDPNP
EVKTPGEKQSQ"
BASE COUNT 232 a 112 c 151 g 111 t
ORIGIN
Query Match 94.1%; Score 506; DB 8; Length 606;
Best Local Similarity 98.5%; Pred. No. 7.7e-116;
Matches 522; Conservative 0; Mismatches 5; Indels 3; Gaps 1;
QY 12 CTGTGAGGAGCTGCTGTGTGTTCTCGCGTCCGAGCTCTTTTCTCTACTGAGATCA 71
Db 1 CTGTGAGGAGCTGCTGTGTGTTCTCGCGTCCGAGCTCTTTTCTCTACTGAGATCA 60
QY 72 TCTGTGTAATATGAGTTGGGCTATGGGCGCCGAGCAAGATCGAGTCAAGAGCGC 128
Db 61 TCTGTGTAATATGAGTTGGGCTATGGGCGCCGAGCAAGATCGAGTCAAGAGCGC 120
QY 129 TACGTAGAGCTCTGAAATGATTTGGGCTATGGGCGCCGAGCAAGATCGAGTCAAGAGTG 188
Db 121 TATGTACAGGCTCTGAAATGATTTGGGCTATGGGCGCCGAGCAAGATCGAGTCAAGAGTG 180
QY 189 GAACACGACACCTGAAAGAGGGAACACGCAACTCAACGTCAGGATCCTCGAGTCT 248
Db 181 GAACACGACACCTGAAAGAGGGAACACGCAACTCAACGTCAGGATCCTCGAGTCT 240
QY 249 CAGGAGGAGGAGTATGAGGAGCATCTCGAGGTCAAGGGCCGAGCAAGCTCAATAGC 308
|||||

Db 241 CAGGAGGAGAGGATGAGGAGCATCTGCAGTCAAGGCCGAGCCTGAAGCTCATAGC 300
Qy 309 CAGGAACAGGGTCAACACAGACTGGGTGTGAGTGTGAAGATGGTCTCTGATGGCGAGGAG 368
Db 301 CAGGAACAGGGTCAACACAGACTGGGTGTGAGTGTGAAGATGGTCTCTGATGGCGAGGAG 360
Qy 369 ATGACCCGCCCAATCCAGAGGAGTGAAGACGCTGAGAGAGGTGAAGCAATCACAG 428
Db 361 ATGACCCGCCCAATCCAGAGGAGGTGAAGACGCTGAGAGAGGTGAAGCAATCACAG 420
Qy 429 TGTAAAGAGACACAGTGTGAATGATGAGGCTGCTCTCTATGTTGGAATTTGTTCAAT 488
Db 421 TGTAAAGAGAGGACGCTGTGAATGATGAGGCTGCTCTCTATGTTGGAATTTGTTCAAT 480
Qy 489 AAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 538
Db 481 AAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 530

RESULT 13
LOCUS AR028491 532 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 17 from patent US 5858689.
ACCESSION AR028491
VERSION AR028491.1 GI:5940464
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 532)
AUTHORS van der Bruggen,P., van den Eynde,B., DeBacker,O. and Boon-Falleur,T.
TITLE Isolated peptides derived from the gage tumor rejection antigen precursor and uses thereof
JOURNAL Patent: US 5858689-A 17 12-JAN-1999;
FEATURES Location/Qualifiers
source 1..532
/organism="unknown"

BASE COUNT 156 a 111 c 154 g 111 t
ORIGIN

Query Match 93.8%; Score 504.8; DB 5; Length 532;
Best Local Similarity 98.1%; Pred. No. 1.5e-115;
Matches 522; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

Qy 10 AGCTGTGAGGAGTGTGTGTTCTCTGCGTCCGAGCTCTTTTCTCTACTGAGATT 69
Db 1 AGCTGTGAGGAGTGTGTGTTCTCTGCGTCCGAGCTCTTTTCTCTACTGAGATT 60

Qy 70 CATCTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCTTAGACCAAGAC 126
Db 61 CATCTGTGAAATATGAGTTGGCGAGGAAGATCGACCCTATTATTGGCCTAGACCAAGGC 120

Qy 127 GCTACGTAGAGCCTCTGAAATGATTGGGCCCTATGCGGCCGAGCAGTTCAAGTGAAG 186
Db 121 GCTATGTGAAATATGAGTTGGCGAGGAAGATCGACCCTATTATTGGCCTAGACCAAGGC 120

Qy 187 TGGACCAAGCAACCTGAAAGAGGGGAACAGCAACTCAACGTCAGGATCTCGAGCTG 246
Db 181 TGGAAACCAAGCAACCTGAAAGAGGGGAACAGCAACTCAACGTCAGGATCTCGAGCTG 240

Qy 247 CTCAGGAGGAGAGGATGAGGAGCATCTGCAGGTCAAGGCCGAGCCTGAAGCTCAT 306
Db 241 CTCAGGAGGAGAGGATGAGGAGCATCTGCAGGTCAAGGCCGAGCCTGAAGCTCAT 300

Qy 307 GCCAGGACAGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTCTGATGGCGAGG 366
Db 301 GCCAGGAAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTCTGATGGCGAGG 360

Qy 367 AGATGGACCCGCCCAATCCAGAGGAGTGAAGAACGCTCAAGGCCGAGCCTGAAGCTCAT 426
Db 361 AGATGGACCCGCCCAATCCAGAGGAGTGAAGAACGCTCAAGGCCGAGCCTGAAGCTCAT 420

Qy 427 AGTGTAAAGAGAGACAGCTTGAATGATGAGGCTCTCTCTATGTTGGAATTTGTTCA 486
Db 421 AGTGTAAAGAGAGGACAGCTTGAATGATGAGGCTCTCTCTATGTTGGAATTTGTTCA 480

Qy 487 TTAATAATCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 538
Db 481 TTAATAATCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 532

RESULT 15
AX334151

Qy 427 AGTGTAAAGAGAGACACGTTGAAATGATGAGGCTGCTCTCTATGTTGGAATTTGTTCA 486
Db 421 AGTGTAAAGAGAGGACAGCTTGAATGATGAGGCTGCTCTCTATGTTGGAATTTGTTCA 480

Qy 487 TTAATAATCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 538
Db 481 TTAATAATCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 532

RESULT 14
LOCUS I55854 532 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 17 from patent US 5648226.
ACCESSION I55854
VERSION I55854.1 GI:2476648
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 532)
AUTHORS Van den Eynde,B., DeBacker,O. and Boon-Falleur,T.
TITLE Isolated peptides derived from tumor rejection antigens, and their use
JOURNAL Patent: US 5648226-A 17 15-JUL-1997;
FEATURES Location/Qualifiers
source 1..532
/organism="unknown"

BASE COUNT 156 a 111 c 154 g 111 t
ORIGIN

Query Match 93.8%; Score 504.8; DB 5; Length 532;
Best Local Similarity 98.1%; Pred. No. 1.5e-115;
Matches 522; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

Qy 10 AGCTGTGAGGAGTGTGTGTTCTCTGCGTCCGAGCTCTTTTCTCTACTGAGATT 69
Db 1 AGCTGTGAGGAGTGTGTGTTCTCTGCGTCCGAGCTCTTTTCTCTACTGAGATT 60

Qy 70 CATCTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCTTAGACCAAGAC 126
Db 61 CATCTGTGAAATATGAGTTGGCGAGGAAGATCGACCCTATTATTGGCCTAGACCAAGGC 120

Qy 127 GCTACGTAGAGCCTCTGAAATGATTGGGCCCTATGCGGCCGAGCAGTTCAAGTGAAG 186
Db 121 GCTATGTACAGCCTCTGAAAGTATTGGGCCCTATGCGGCCGAGCAGTTCAAGTGAAG 180

Qy 187 TGGAAACCAAGCAACCTGAAAGAGGGGAACAGCAACTCAACGTCAGGATCTCGAGCTG 246
Db 181 TGGAAACCAAGCAACCTGAAAGAGGGGAACAGCAACTCAACGTCAGGATCTCGAGCTG 240

Qy 247 CTCAGGAGGAGAGGATGAGGAGCATCTGCAGGTCAAGGCCGAGCCTGAAGCTCAT 306
Db 241 CTCAGGAGGAGAGGATGAGGAGCATCTGCAGGTCAAGGCCGAGCCTGAAGCTCAT 300

Qy 307 GCCAGGACAGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTCTGATGGCGAGG 366
Db 301 GCCAGGAAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTCTGATGGCGAGG 360

Qy 367 AGATGGACCCGCCCAATCCAGAGGAGTGAAGAACGCTCAAGGCCGAGCCTGAAGCTCAT 426
Db 361 AGATGGACCCGCCCAATCCAGAGGAGTGAAGAACGCTCAAGGCCGAGCCTGAAGCTCAT 420

Qy 427 AGTGTAAAGAGAGACAGCTTGAATGATGAGGCTCTCTCTATGTTGGAATTTGTTCA 486
Db 421 AGTGTAAAGAGAGGACAGCTTGAATGATGAGGCTCTCTCTATGTTGGAATTTGTTCA 480

Qy 487 TTAATAATCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 538
Db 481 TTAATAATCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 532

RESULT 15
AX334151

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2003, 12:41:59 ; Search time 1040 Seconds
(without alignments)
8335.716 Million cell updates/sec

Title: US-09-782-745-14
Perfect score: 538
Sequence: 1 ACGCAGGAGCTGTGAGGC.....CTGCAAGAAAAA 538

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 60155 seqs, 8056825618 residues

Total number of hits satisfying chosen parameters: 120310

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl HTG:* *

1: gb_htg:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	163.2	30.3	240000	1 AC009528	Homo sapi
C 2	140.8	26.2	190309	1 AL645949	Homo sapi
C 3	131.4	24.4	48802	1 AC068431	Homo sapi
C 4	118	21.9	110000	1 AL831785	Homo sapi
C 5	97.2	18.1	48802	1 AC068431	Homo sapi
C 6	69.2	12.9	81953	1 AC016835	Homo sapi
C 7	69.2	12.9	93319	1 AC002415	Homo sapi
C 8	69.2	12.9	114344	1 AL445227	Homo sapi
C 9	69.2	12.9	114344	1 AL445227	Homo sapi
C 10	67.8	12.6	171489	1 AL772246	Homo sapi
C 11	67.6	12.6	180859	1 AC025553	Homo sapi
C 12	66	12.3	171489	1 AL772246	Homo sapi
C 13	64.6	12.0	180859	1 AC025553	Homo sapi
C 14	62.6	11.6	64891	1 AC118662	Homo sapi
C 15	60.6	11.3	93319	1 AC002415	Homo sapi
C 16	54.2	10.1	185032	1 AC109034	Rattus no
C 17	53.4	9.9	231755	1 AC115480	Rattus no
C 18	53.2	9.5	198806	1 AL671904	Mus muscu
C 19	50	9.3	33082	1 AC091365	Rattus no
C 20	49.6	9.2	188082	1 AC121784	Mus muscu
C 21	49.6	9.2	214690	1 AC083889	Mus muscu
C 22	49.6	9.2	221048	1 AC087329	Mus muscu
C 23	49.2	9.1	157392	1 AC120934	Rattus no
C 24	49.2	9.1	165210	1 AC101542	Mus muscu
C 25	49.2	9.1	196420	1 AC108825	Mus muscu
C 26	49	9.1	36077	1 AC106641	Rattus no
C 27	49	9.1	87750	1 AC111854	Rattus no
C 28	49	9.1	110141	1 AC114072	Rattus no
C 29	48.8	9.1	139655	1 AC125152	Mus muscu
C 30	48.6	9.0	110558	1 AC115492	Rattus no
C 31	48.2	9.0	63239	1 AC112945	Mus muscu
C 32	48	8.9	51267	1 AC100027	Mus muscu

C 33	48	8.9	165042	1 AC115567	AC115567 Rattus no
C 34	48	8.9	166299	1 AC127244	AC127244 Mus muscu
C 35	48	8.9	166299	1 AC127244	AC127244 Mus muscu
C 36	48	8.9	187008	1 AC125219	AC125219 Mus muscu
C 37	48	8.9	205802	1 AC123922	AC123922 Mus muscu
C 38	48	8.9	210727	1 AC122864	AC122864 Mus muscu
C 39	48	8.9	222259	1 AC113078	AC113078 Mus muscu
C 40	48	8.9	226997	1 AC128798	AC128798 Rattus no
C 41	47.8	8.9	162987	1 AC101987	AC101987 Mus muscu
C 42	47.8	8.9	211281	1 AL713985	AL713985 Mus muscu
C 43	47.6	8.8	173732	1 AC099625	AC099625 Mus muscu
C 44	47.4	8.8	186076	1 AC127664	AC127664 Rattus no
C 45	47.4	8.8	186398	1 AC098110	AC098110 Rattus no

ALIGNMENTS

RESULT 1
AC009528/c
LOCUS
DEFINITION Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 21
AC009528 240000 bp DNA linear HTG 19-DEC-1999
unorderd pieces.

AC009528 7 GI:6604463
VERSION
KEYWORDS HTG; HTGS PHASE1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL 1 (bases 1 to 240000)
REFERENCE Direct Submission
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
JOURNAL 2 (bases 1 to 240000)
REFERENCE Direct Submission
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
JOURNAL Submitted (26-AUG-1999) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
On Dec 20, 1999 this sequence version replaced gi:6573827.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1	1416:	contig of 1416 bp in length
* 1417	1715:	gap of unknown length
* 1716	3005:	contig of 1290 bp in length
* 3006	3304:	gap of unknown length
* 3305	4379:	contig of 1075 bp in length
* 4380	4678:	gap of unknown length
* 4679	6718:	contig of 2040 bp in length
* 6719	7017:	gap of unknown length
* 7018	8520:	contig of 1503 bp in length
* 8521	8819:	gap of unknown length
* 8820	10657:	contig of 1838 bp in length
* 10658	10956:	gap of unknown length
* 10957	13220:	contig of 2264 bp in length
* 13221	13519:	gap of unknown length
* 13520	15227:	contig of 1708 bp in length
* 15228	15526:	gap of unknown length
* 15527	18339:	contig of 2613 bp in length
* 18140	18438:	gap of unknown length
* 18439	20233:	contig of 1795 bp in length
* 20234	20532:	gap of unknown length
* 20533	22806:	contig of 2274 bp in length
* 22807	23104:	gap of unknown length
* 23105	24924:	contig of 1820 bp in length
* 24925	25222:	gap of unknown length
* 25223	27314:	contig of 2092 bp in length

27315 27612: gap of unknown length
* 27613: contig of 4564 bp in length
* 32177 32176: gap of unknown length
* 32178 32177: gap of unknown length
* 32475 43341: contig of 10867 bp in length
* 43342 43341: contig of unknown length
* 43640 65232: contig of 21653 bp in length
* 65293 65590: gap of unknown length
* 65591 90191: contig of 24601 bp in length
* 90192 90489: gap of unknown length
* 90490 115831: contig of 25342 bp in length
* 115832 116129: gap of unknown length
* 116130 143584: contig of 27455 bp in length
* 143585 143882: gap of unknown length
* 143883 175459: contig of 31577 bp in length
* 175460 175757: gap of unknown length
* 175758 240000: contig of 64243 bp in length.

FEATURES
source
1. .240000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"

BASE COUNT 61984 a 55073 c 54778 g 62188 t 5977 others
ORIGIN

Query Match 30.3%; Score 163.2; DB 1; Length 240000;
Best Local Similarity 70.0%; Pred. No. 1.3e-26;
Matches 346; Conservative 0; Mismatches 113; Indels 35; Gaps 8;

QY 1 AGCCAGGGAGCTGTGAGCAGTCTGTGTGTTCTCTCGCGTCCGACCTTTTTCCTCT 60
Db 99869 AAGAAGTGTCTCGGGAGCTGTGAGGATATGCTGTACATTCGCGCGGTGAATCTCTT 99810

QY 61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGCCCTAGAC 120
Db 99809 ACTGAGACTCATCTGTGTGAA--ATGAGTCGTTGAGGAGATCA----ATCAGCCTCGAC 99756

QY 121 CAAGACGCTACGTAGAGCTCTCTGAAATGATTGGGCTATGCGGCCCGAGCAGTTCAGTG 180
Db 99755 CAATACAACTGTACAACTTAAGG---TGATCTGGCTATTCTGCCAAGCAGTTCAGTC 99699

QY 181 ATGAGTGGNACAGCAACACCTTAAGNAGGGGACACGACCTCAACCTCAGGATCCTG 240
Db 99698 ATGAAC-----CACCTGAAGAAGTGGAAACCACTCAAGTCAAGATCTT 99651

QY 241 CAGCTGCTCAGGAGGAGGATGAGGAGCATCTGCAGGTCATCGAGGTCGAAAGCCGAAAG 300
Db 99650 CACCTGCTCAGGAGAGAGAGAGCAGGAGCATCTGAGCTCAAGCCCTGAGCTGAAG 99591

QY 301 CTCATAGCCAGGAACAGGGTCAACCCACAGACTGGGTGTGAGTGTGAAGATGTCCTGATG 360
Db 99590 CTGATTCGCGAGGAACTGGTT---CAAAGACTGGGTATGAGCATGAGATGATCCTGATG 99535

QY 361 GCAGGAGATGGACCCGCAATCCAGAGGAGTGAACCCCTGAAGAGTGAAGAGC 420
Db 99534 TGAAGGGATGTGCCTACTTAACCCCGAGAGCTGGGAACCTGC--CAGAAGGTGAAGAGC 99477

QY 421 AATCACAGTGTAAAGAAGACAGCTTGAATGATGACAGGCTGCTCCTATGTTGAAAT 480
Db 99476 ATACAGCT-----TAAAGACATGCTGAATGATGCGGGCTG---CTATGTTGGAAT 99425

QY 481 TGTTCATTAATAATT 494
Db 99424 TGTTCATTAATAAGT 99411

RESULT 2
AL645949/c 190309 bp DNA linear HTG 16-AUG-2002
LOCUS Homo sapiens chromosome 4 clone RP11-357G3, *** SEQUENCING IN
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AL645949
VERSION AL645949.11 GI:22316144
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 190309)
Ashwell,R.
REFERENCE Direct Submission
AUTHORS Submitted (15-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
TITLE Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
JOURNAL humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Aug 19, 2002 this sequence version replaced gi:21953022.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA357G3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 2% of reads
Sequencing vector: plasmid; L08752; 97% of reads
Chemistry: Dye-terminator ABI; 0% of reads
Chemistry: Dye-terminator; 4% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 95% of reads
Consensus quality: 190103 bases at least Q40
Consensus quality: 190158 bases at least Q30
Consensus quality: 190173 bases at least Q20
Insert size: 190309; sum-of-contigs
Insert size: 189393; 6.6% error; agarose-fp
Quality coverage: 7.95x in Q20 bases; sum-of-contigs Quality
coverage: 8.10x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1. 190309
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-357G3"
/clone.lib="RP11-11.2"
misc_feature 1. 190309
/note="assembly fragment:04435"

BASE COUNT 50017 a 50895 c 46222 g 43175 t
ORIGIN

Query Match 26.2%; Score 140.8; DB 1; Length 190309;
Best Local Similarity 64.6%; Pred. No. 1.2e-21;
Matches 331; Conservative 0; Mismatches 152; Indels 29; Gaps 7;

QY 1 AGCCAGGGAGCTGTGAGGAGTGTGTGTG-GTTCCTGCGTCCGAGCTCTTTTCTCTC 59
Db 102418 ACACGTGACCTGTGAGGCTGTGAGGTTGCGTTCCTGCTGAGACTCTTTTACCCAC 102359

QY 60 TACTGAGATTCATCTGTGAAATATGAGTTGGCAGGAAGATCGACCTATCGGCCTAGA 119
Db 102358 TACTGAGAAGCAGCCGTTTGAATATGAGTTGGCAAGGAAGATCAGCATGTAGGCTTGA 102299

QY 120 CCAAGACGCTACGTAGAGCTCTCTGAAATGATTGGGCTATGCGGCCCGGAGCAGTTCAGT 179
Db 102298 CCAAGACGCTATGTGAGGCTCTCTGAGCTAACTGGGCTGTGC-----TTGAGCCCACT 102245

QY 180 GATGAAGTGAACACCAACACCTGAAGAGGGGAAACGACAACTCAACGTCAGGATCCT 239
Db 102244 GATGAGCAG-----CCTCAGCAAGAGGAAACCAACCACTTGAAGTCGGGCTCTT 102197

QY 240 GCAGCTGTCTCAGGAGGA-GAGGATGAGGAGGAGCATCTGACGATCAAGGCGGAGCCTGA 298
Db 102196 ACCTGGCCAGGAGAGAGGAAGATCACCGTGTGCTGCTGAGATTCCTTGTGCTTGACCGGA 102137

* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1	715:	contig of 715 bp in length
716	815:	gap of 100 bp
816	1527:	contig of 712 bp in length
1528	1627:	gap of 100 bp
1628	2361:	contig of 734 bp in length
2362	2461:	gap of 100 bp
2462	3178:	contig of 717 bp in length
3179	3278:	gap of 100 bp
3279	4003:	contig of 725 bp in length
4004	4103:	gap of 100 bp
4104	4836:	contig of 733 bp in length
4837	4936:	gap of 100 bp
4937	5560:	contig of 724 bp in length
5561	5760:	gap of 100 bp
5761	6486:	contig of 726 bp in length
6487	6586:	gap of 100 bp
6587	7310:	contig of 724 bp in length
7311	7410:	gap of 100 bp
7411	8116:	contig of 706 bp in length
8117	8216:	gap of 100 bp
8217	8947:	contig of 731 bp in length
8948	9047:	gap of 100 bp
9048	9758:	contig of 711 bp in length
9759	9858:	gap of 100 bp
9859	10589:	contig of 731 bp in length
10590	10689:	gap of 100 bp
10690	11412:	contig of 723 bp in length
11413	11512:	gap of 100 bp
11513	12237:	contig of 725 bp in length
12238	12337:	gap of 100 bp
12338	13067:	contig of 730 bp in length
13068	13167:	gap of 100 bp
13168	13906:	contig of 739 bp in length
13907	14006:	gap of 100 bp
14007	14718:	contig of 712 bp in length
14719	14818:	gap of 100 bp
14819	15543:	contig of 725 bp in length
15544	15643:	gap of 100 bp
15644	16372:	contig of 729 bp in length
16373	16472:	gap of 100 bp
16473	17215:	contig of 743 bp in length
17216	17315:	gap of 100 bp
17316	18046:	contig of 731 bp in length
18047	18146:	gap of 100 bp
18147	18877:	contig of 731 bp in length
18878	18977:	gap of 100 bp
18979	19695:	contig of 718 bp in length
19696	19795:	gap of 100 bp
19796	20524:	contig of 729 bp in length
20525	20624:	gap of 100 bp
20625	21366:	contig of 742 bp in length
21367	21466:	gap of 100 bp
21467	22181:	contig of 715 bp in length
22182	22381:	gap of 100 bp
22382	22996:	contig of 715 bp in length
22997	23096:	gap of 100 bp
23097	23823:	contig of 727 bp in length
23824	23923:	gap of 100 bp
23924	24655:	contig of 732 bp in length
24656	24755:	gap of 100 bp
24756	25496:	contig of 741 bp in length
25497	25596:	gap of 100 bp
25597	26345:	contig of 749 bp in length
26346	26445:	gap of 100 bp

QY	299	AGCTCATAGCCAGGAAACAGGGTACCCACAGACTGGGTGTAGTGTGAAGATGGTCTCTGA	358		
Db	102136	AGCTGATCTCCGGGAGCTGTCTCAAAAGAC-----TGGGATGAATGTGGAGATGGTCTCTGA	102082		
QY	359	TGGCGAGGAGATGGACCGCCCAATCCAGAGGAGGTGAACAGCCCTGAAGAAGGTGAAAA	418		
Db	102081	T-GTCCGGGGATATTCTGCCGAAATCAGAGCAATTTAAATTTGCCAGAAGGAGGTGAAGG	102023		
QY	419	GCAATCACACTGTATAAAGAGACACGTTGAATATGATGACAGCTCTCTTATGTTGGAAA	478		
Db	102022	GCAACACAGGTTCCAGAGGAGACAAGCTGAAACAATGCAAACTGGTTTATTATTAGATA	101963		
QY	479	TTTGTTCATTAAAAATTTCTCCCAATAAGCTTT	510		
Db	101962	CGTG--ACTTAAAAATATCTCAATACAGTTTT	101934		
RESULT 3					
LOCUS	AC068431	48802 bp	DNA linear HTG 02-MAY-2000		
DEFINITION	Homo sapiens chromosome 1 clone RP11-211H10 map 1, LOW-PASS SEQUENCE SAMPLING.				
ACCESSION	AC068431				
VERSION	AC068431.1	GI: 67677847			
KEYWORDS	HTG; HTGS; PHASE0.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini, Hominidae; Homo.				
AUTHORS	1 (bases 1 to 48802)				
TITLE	Birren, B., Linton, L., Nusbaum, C. and Lander, E.				
JOURNAL	Homo sapiens chromosome 1, clone RP11-211H10				
REFERENCE	2 (bases 1 to 48802)				
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F., Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Headford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Menga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Ollivier, T. M., Oliver, J., Peterson, K., Pierre, N., Pisanu, C., Pollard, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tassfay, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT	All repeats were identified using RepeatMasker: Smit, A. F. A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html				
	----- Genome Center				
	Center: Whitehead Institute/ MIT Center for Genome Research				
	Center code: WIBR				
	Web site: http://www-seq.wi.mit.edu				
	Contact: sequence_submissions@genome.wi.mit.edu				
	----- Project Information				
	Center project name: L7691				
	Center clone name: 211_H_10				

* NOTE: This record contains 59 individual					
* sequencing reads that have not been assembled into					

* 27549 29549: contig of 2001 bp in length
* 29550 29649: gap of 100 bp
* 32896 32896: contig of 3247 bp in length
* 32897 32996: gap of 100 bp
* 32937 32977: contig of 3281 bp in length
* 36278 36377: gap of 100 bp
* 36378 38755: contig of 2378 bp in length
* 38756 38855: gap of 100 bp
* 38856 41582: contig of 2727 bp in length
* 41583 41682: gap of 100 bp
* 41683 44651: contig of 2969 bp in length
* 44652 44751: gap of 100 bp
* 44752 47834: contig of 3083 bp in length
* 47835 47934: gap of 100 bp
* 47935 59046: contig of 11112 bp in length
* 59047 59146: gap of 100 bp
* 59147 61468: contig of 2322 bp in length
* 61469 61568: gap of 100 bp
* 61569 63827: contig of 2259 bp in length
* 63828 63927: gap of 100 bp
* 63928 67325: contig of 3398 bp in length
* 67326 67425: gap of 100 bp
* 67426 69741: contig of 2316 bp in length
* 69742 69841: gap of 100 bp
* 69842 72033: contig of 2192 bp in length
* 72034 72133: gap of 100 bp
* 72134 74258: contig of 2125 bp in length
* 74259 74358: gap of 100 bp
* 74359 77078: contig of 2720 bp in length
* 77079 77178: gap of 100 bp
* 77179 80648: contig of 3470 bp in length
* 80649 80748: gap of 100 bp
* 80749 84081: contig of 3333 bp in length
* 84082 84181: gap of 100 bp
* 84182 87056: contig of 2875 bp in length
* 87057 87156: gap of 100 bp
* 87157 89197: contig of 2041 bp in length
* 89198 89297: gap of 100 bp
* 89298 91645: contig of 2348 bp in length
* 91646 91745: gap of 100 bp
* 91746 94371: contig of 2626 bp in length
* 94372 94471: gap of 100 bp
* 94472 96563: contig of 2092 bp in length
* 96564 96663: gap of 100 bp
* 96664 100313: contig of 3650 bp in length
* 100314 100413: gap of 100 bp
* 100414 102923: contig of 2510 bp in length
* 102924 103023: gap of 100 bp
* 103024 105754: contig of 2731 bp in length
* 105755 105854: gap of 100 bp
* 105855 108005: contig of 2151 bp in length
* 108006 108105: gap of 100 bp
* 108106 113234: contig of 5129 bp in length
* 113235 113334: gap of 100 bp
* 113335 115344: contig of 2010 bp in length
* 115345 115444: gap of 100 bp
* 115445 120013: contig of 4569 bp in length
* 120014 120113: gap of 100 bp
* 120114 122698: contig of 2585 bp in length
* 122699 122798: gap of 100 bp
* 122799 127907: contig of 5109 bp in length
* 127908 128007: gap of 100 bp
* 128008 130666: contig of 2659 bp in length
* 130667 130766: gap of 100 bp
* 130767 133234: contig of 2468 bp in length
* 133235 133334: gap of 100 bp
* 133335 136699: contig of 3365 bp in length
* 136700 136799: gap of 100 bp
* 136800 139189: contig of 2390 bp in length
* 139190 139289: gap of 100 bp
* 139290 142287: contig of 2998 bp in length
* 142288 142387: gap of 100 bp
* 142388 149385: contig of 6998 bp in length

* 149386 149485: gap of 100 bp
* 149486 151937: contig of 2452 bp in length
* 151938 152037: gap of 100 bp
* 152038 154126: contig of 2089 bp in length
* 154127 154226: gap of 100 bp
* 154227 157753: contig of 3527 bp in length
* 157754 157853: gap of 100 bp
* 157854 159927: contig of 2074 bp in length
* 159928 160027: gap of 100 bp
* 160028 162267: contig of 2240 bp in length
* 162268 162367: gap of 100 bp
* 162368 167229: contig of 4862 bp in length
* 167230 167329: gap of 100 bp
* 167330 169663: contig of 2334 bp in length
* 169664 169763: gap of 100 bp
* 169764 172099: contig of 2336 bp in length
* 172100 172199: gap of 100 bp
* 172200 175654: contig of 3455 bp in length
* 175655 175754: gap of 100 bp
* 175755 180193: contig of 4439 bp in length
* 180194 180293: gap of 100 bp
* 180294 186167: contig of 5874 bp in length
* 186168 186267: gap of 100 bp
* 186268 188572: contig of 2305 bp in length
* 188573 188672: gap of 100 bp
* 188673 191418: contig of 2746 bp in length
* 191419 191518: gap of 100 bp
* 191519 200088: contig of 8570 bp in length
* 200089 200188: gap of 100 bp
* 200189 205237: contig of 5049 bp in length
* 205238 205337: gap of 100 bp
* 205338 208097: contig of 2760 bp in length
* 208098 208197: gap of 100 bp
* 208198 211472: contig of 3275 bp in length
* 211473 211572: gap of 100 bp
* 211573 213769: contig of 2197 bp in length
* 213770 213869: gap of 100 bp
* 213870 216122: contig of 2253 bp in length
* 216123 216222: gap of 100 bp
* 216223 219243: contig of 3021 bp in length
* 219244 219343: gap of 100 bp
* 219344 224613: contig of 5270 bp in length
* 224614 224713: gap of 100 bp
* 224714 226811: contig of 2098 bp in length
* 226812 226911: gap of 100 bp
* 226912 230682: contig of 3771 bp in length
* 230683 230782: gap of 100 bp
* 230783 232928: contig of 2145 bp in length
* 232928 233027: gap of 100 bp
* 233028 235051: contig of 2024 bp in length
* 235052 235151: gap of 100 bp
* 235152 237642: contig of 2491 bp in length
* 237643 237742: gap of 100 bp
* 237743 240388: contig of 2646 bp in length
* 240389 240488: gap of 100 bp
* 240489 242671: contig of 2183 bp in length
* 242672 242771: gap of 100 bp
* 242772 245337: contig of 2566 bp in length
* 245338 245437: gap of 100 bp
* 245438 248765: contig of 3328 bp in length
* 248766 248865: gap of 100 bp
* 248866 250924: contig of 2059 bp in length
* 250925 251024: gap of 100 bp
* 251025 253405: contig of 2381 bp in length

Query Match 21.9%; Score 118; DB 1; Length 110000;
Best Local Similarity 96.0%; Pred. No. 1.4e-16;
Matches 121; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 403 CTGAAGAGGTGAAAAGCAATCACAGTGTAAAAAGACACAGCTTGAATGATGCAGGCT 462
Db 67467 CTGTACAGGTGAAAAGCAATCACAGTGTAAAAAGAGGCAGCTTGAATGATGCAGGCT 67526

```

QY 463 GCTCCTATGTTGGAATTTGTTCAATAAATTCCTCCCAATAAGCTTTACAGCCTTCGCG 522
Db 67527 GCTCCTATGTTGGAATTTGTTCAATAAATTCCTCCCAATAAGCTTTACAGCCTTCGCG 67586

QY 523 AAAGAA 528
Db 67587 AAAGAA 67592
|||||

RESULT 5
AC068431/c
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-211H10 map 1, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC068431
VERSION 1 GI:7677847
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 48802)
Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castie, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliiev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lenoczky, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, N., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7691
Center Clone name: 211_H_10
-----
* NOTE: This record contains 59 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

```

1 715: contig of 715 bp in length
716 815: gap of 100 bp
816 1527: contig of 712 bp in length
1528 1627: gap of 100 bp
1628 2361: contig of 734 bp in length
2362 2461: gap of 100 bp
2462 3178: contig of 717 bp in length
3179 3278: gap of 100 bp
3279 4003: contig of 725 bp in length
4004 4103: gap of 100 bp
4104 4836: contig of 733 bp in length
4837 4936: gap of 100 bp
4937 5660: contig of 724 bp in length
5661 5760: gap of 100 bp
5761 6486: contig of 726 bp in length
6487 6586: gap of 100 bp
6587 7310: contig of 724 bp in length
7311 7410: gap of 100 bp
7411 8116: contig of 706 bp in length
8117 8216: gap of 100 bp
8217 8947: contig of 731 bp in length
8948 9047: gap of 100 bp
9048 9758: contig of 711 bp in length
9759 9858: gap of 100 bp
9859 10589: contig of 731 bp in length
10590 10689: gap of 100 bp
10690 11412: contig of 723 bp in length
11413 11512: gap of 100 bp
11513 12237: contig of 725 bp in length
12238 12337: gap of 100 bp
12338 13067: contig of 730 bp in length
13068 13167: gap of 100 bp
13168 13906: contig of 739 bp in length
13907 14066: gap of 100 bp
14067 14718: contig of 712 bp in length
14719 14818: gap of 100 bp
14819 15543: contig of 725 bp in length
15544 15643: gap of 100 bp
15644 16372: contig of 729 bp in length
16373 16472: gap of 100 bp
16473 17215: contig of 743 bp in length
17216 17315: gap of 100 bp
17316 18046: contig of 731 bp in length
18047 18146: gap of 100 bp
18147 18877: contig of 731 bp in length
18878 18977: gap of 100 bp
18978 19695: contig of 718 bp in length
19696 19795: gap of 100 bp
19796 20524: contig of 729 bp in length
20525 20624: gap of 100 bp
20625 21366: contig of 742 bp in length
21367 21466: gap of 100 bp
21467 22181: contig of 715 bp in length
22182 22281: gap of 100 bp
22282 22996: contig of 715 bp in length
22997 23096: gap of 100 bp
23097 23823: contig of 727 bp in length
23824 23923: gap of 100 bp
23924 24655: contig of 732 bp in length
24656 24755: gap of 100 bp
24756 25496: contig of 741 bp in length
25497 25596: gap of 100 bp
25597 26345: contig of 749 bp in length
26346 26445: gap of 100 bp
26446 27190: contig of 745 bp in length
27191 27290: gap of 100 bp
27291 28031: contig of 741 bp in length
28032 28131: gap of 100 bp
28132 28855: contig of 724 bp in length
28856 28955: gap of 100 bp
28956 29676: contig of 721 bp in length
29677 29776: gap of 100 bp
29777 30506: contig of 730 bp in length

```

```

* 30507 30606: gap of 100 bp
* 30607 31326: contig of 720 bp in length
* 31327 31426: gap of 100 bp
* 31427 32153: contig of 727 bp in length
* 32154 32253: gap of 100 bp
* 32254 32982: contig of 729 bp in length
* 32983 33082: gap of 100 bp
* 33083 33798: contig of 716 bp in length
* 33799 33898: gap of 100 bp
* 33899 34630: contig of 732 bp in length
* 34631 34730: gap of 100 bp
* 34731 35456: contig of 726 bp in length
* 35457 35556: gap of 100 bp
* 35557 36283: contig of 727 bp in length
* 36284 36383: gap of 100 bp
* 36384 37127: contig of 744 bp in length
* 37128 37227: gap of 100 bp
* 37228 37945: contig of 718 bp in length
* 37946 38045: gap of 100 bp
* 38046 38775: contig of 730 bp in length
* 38776 38875: gap of 100 bp
* 38876 39618: contig of 743 bp in length
* 39619 39718: gap of 100 bp
* 39719 40451: contig of 733 bp in length
* 40452 40551: gap of 100 bp
* 40552 41290: contig of 739 bp in length
* 41291 41390: gap of 100 bp
* 41391 42121: contig of 731 bp in length
* 42122 42221: gap of 100 bp
* 42222 42938: contig of 717 bp in length
* 42939 43038: gap of 100 bp
* 43039 43771: contig of 733 bp in length
* 43772 43871: gap of 100 bp
* 43872 44617: contig of 746 bp in length
* 44618 44717: gap of 100 bp
* 44718 45470: contig of 753 bp in length
* 45471 45570: gap of 100 bp
* 45571 46302: contig of 732 bp in length
* 46303 46402: gap of 100 bp
* 46403 47129: contig of 727 bp in length
* 47130 47229: gap of 100 bp
* 47230 47966: contig of 737 bp in length
* 47967 48066: gap of 100 bp
* 48067 48802: contig of 736 bp in length.

```

FEATURES

```

source
1. .48802
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosomes="1"
/maps="1"
/clone="RP11-211H10"
/clone_lib="RP11-11 Human Male BAC"

```

```

BASE COUNT 11763 a 9361 c 9024 g 12533 t 6121 others
ORIGIN

```

```

Query Match 18.1%; Score 97.2; DB 1; Length 48802;
Best Local Similarity 92.6%; Pred. No. 6.1e-12;
Matches 113; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
QY 281 TCAGGCGCAGCCTGAGCTCATAGCCAGGACAGGTCACCCACAGACTGGGTCTGA 340
|||||
Db 14163 TTAGGCGCCGAGCCTGAGCTCATAGCCAGGACAGGTCACCCACAGACTGGGTCTGA 14104
|||||
QY 341 GTGTGAAGATGGTCTCTGATGGCGAGGAGATGGACCCGCCAAATCCAGAGAGGTGA AAC 400
|||||
Db 14103 GTGTGAACA-CGCCCTCGAGGCGAGGAGATGGACCCGCCAAATCCAGAGAGGTGA AAC 14045
|||||
QY 401 GC 402
|||
Db 14044 GC 14043

```

RESULT 6

```

AC016835
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

AC016835
Homo sapiens clone RP11-1P5, LOW-PASS SEQUENCE SAMPLING.
AC016835
HTG; HTGS_PHASE0.
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 81953)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-1P5
Unpublished
2 (bases 1 to 81953)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Collings,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gage,D.,
Galsgan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (08-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6539374.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Genome Center
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L2520
Center clone name: 1_P_5

* NOTE: This record contains 91 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 769: contig of 769 bp in length
* 770 869: gap of 100 bp
* 870 1691: contig of 822 bp in length
* 1692 1791: gap of 100 bp
* 1792 2590: contig of 799 bp in length
* 2591 2690: gap of 100 bp
* 2691 3502: contig of 812 bp in length
* 3503 3602: gap of 100 bp
* 3603 4403: contig of 801 bp in length
* 4404 4503: gap of 100 bp
* 4504 5271: contig of 768 bp in length
* 5272 5371: gap of 100 bp
* 5372 6160: contig of 789 bp in length
* 6161 6260: gap of 100 bp
* 6261 7048: contig of 788 bp in length
* 7049 7148: gap of 100 bp
* 7149 7945: contig of 797 bp in length

* 7946 8045: gap of 100 bp
* 8046 884: contig of 819 bp in length
* 8865 8964: gap of 100 bp
* 8965 9788: contig of 804 bp in length
* 9769 9869: gap of 100 bp
* 9869 10679: contig of 811 bp in length
* 10680 10779: gap of 100 bp
* 10780 11545: contig of 766 bp in length
* 11546 11645: gap of 100 bp
* 11646 12460: contig of 815 bp in length
* 12461 12560: gap of 100 bp
* 12561 13372: contig of 812 bp in length
* 13373 13472: gap of 100 bp
* 13473 14275: contig of 803 bp in length
* 14276 14375: gap of 100 bp
* 14376 15173: contig of 798 bp in length
* 15174 15273: gap of 100 bp
* 15274 16058: contig of 785 bp in length
* 16059 16158: gap of 100 bp
* 16159 16953: contig of 795 bp in length
* 16954 17053: gap of 100 bp
* 17054 17840: contig of 787 bp in length
* 17841 17940: gap of 100 bp
* 17941 18750: contig of 810 bp in length
* 18751 18850: gap of 100 bp
* 18851 19556: contig of 806 bp in length
* 19557 19756: gap of 100 bp
* 19757 20561: contig of 805 bp in length
* 20562 20661: gap of 100 bp
* 20662 21456: contig of 795 bp in length
* 21457 21556: gap of 100 bp
* 21557 22382: contig of 826 bp in length
* 22383 22482: gap of 100 bp
* 22483 23306: contig of 824 bp in length
* 23307 23406: gap of 100 bp
* 23407 24217: contig of 811 bp in length
* 24218 24317: gap of 100 bp
* 24318 25116: contig of 799 bp in length
* 25117 25216: gap of 100 bp
* 25217 25999: contig of 783 bp in length
* 26000 26099: gap of 100 bp
* 26100 26879: contig of 780 bp in length
* 26880 26979: gap of 100 bp
* 26980 27773: contig of 794 bp in length
* 27774 27873: gap of 100 bp
* 27874 28668: contig of 795 bp in length
* 28669 28768: gap of 100 bp
* 28769 29592: contig of 824 bp in length
* 29593 29692: gap of 100 bp
* 29693 30523: contig of 831 bp in length
* 30524 30623: gap of 100 bp
* 30624 31423: contig of 800 bp in length
* 31424 31523: gap of 100 bp
* 31524 32321: contig of 798 bp in length
* 32322 32421: gap of 100 bp
* 32422 33244: contig of 823 bp in length
* 33245 33344: gap of 100 bp
* 33345 34172: contig of 828 bp in length
* 34173 34272: gap of 100 bp
* 34273 35079: contig of 807 bp in length
* 35080 35179: gap of 100 bp
* 35180 35967: contig of 788 bp in length
* 35968 36067: gap of 100 bp
* 36068 36846: contig of 779 bp in length
* 36847 36946: gap of 100 bp
* 36947 37740: contig of 794 bp in length
* 37741 37840: gap of 100 bp
* 37841 38651: contig of 811 bp in length
* 38652 38751: gap of 100 bp
* 38752 39566: contig of 815 bp in length
* 39567 39666: gap of 100 bp
* 39667 40472: contig of 806 bp in length
* 40473 40572: gap of 100 bp

* 40573 41381: contig of 809 bp in length
* 41382 41481: gap of 100 bp
* 41482 42296: contig of 815 bp in length
* 42297 42396: gap of 100 bp
* 42397 43213: contig of 817 bp in length
* 43214 43313: gap of 100 bp
* 43314 44124: contig of 811 bp in length
* 44125 44224: gap of 100 bp
* 44225 45028: contig of 804 bp in length
* 45029 45128: gap of 100 bp
* 45129 45937: contig of 809 bp in length
* 45938 46037: gap of 100 bp
* 46038 46814: contig of 777 bp in length
* 46815 46914: gap of 100 bp
* 46915 47710: contig of 796 bp in length
* 47711 47810: gap of 100 bp
* 47811 48580: contig of 770 bp in length
* 48581 48680: gap of 100 bp
* 48681 49465: contig of 785 bp in length
* 49466 49565: gap of 100 bp
* 49566 50377: contig of 812 bp in length
* 50378 50477: gap of 100 bp
* 50478 51260: contig of 783 bp in length
* 51261 51360: gap of 100 bp
* 51361 52143: contig of 783 bp in length
* 52144 52243: gap of 100 bp
* 52244 53049: contig of 806 bp in length
* 53050 53149: gap of 100 bp
* 53150 53959: contig of 810 bp in length
* 53960 54059: gap of 100 bp
* 54060 54832: contig of 773 bp in length
* 54833 54932: gap of 100 bp
* 54933 55712: contig of 780 bp in length
* 55713 55812: gap of 100 bp
* 55813 56604: contig of 792 bp in length
* 56605 56704: gap of 100 bp
* 56705 57487: contig of 783 bp in length
* 57488 57587: gap of 100 bp
* 57588 58395: contig of 808 bp in length
* 58396 58495: gap of 100 bp
* 58496 59302: contig of 807 bp in length
* 59303 59402: gap of 100 bp
* 59403 60205: contig of 803 bp in length
* 60206 60305: gap of 100 bp
* 60306 61119: contig of 814 bp in length
* 61120 61219: gap of 100 bp
* 61220 62026: contig of 807 bp in length
* 62027 62126: gap of 100 bp
* 62127 62941: contig of 815 bp in length
* 62942 63041: gap of 100 bp
* 63042 63857: contig of 816 bp in length
* 63858 63957: gap of 100 bp
* 63958 64772: contig of 815 bp in length
* 64773 64872: gap of 100 bp
* 64873 65675: contig of 803 bp in length

Query Match 12.9% Score 69.2; DB 1; Length 81953;
Best Local Similarity 70.8%; Pred. No. 9.9e-06;
Matches 52; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 284 AGGCCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTACCCACAGACTGGGTGAGTG 343
DB 47098 AGTGCCTTAACCTGGAAGCTGATCTCCAGGAGCTGTCTCAGTCAAGACTGGGATGAATG 47157
QY 344 TGAAGATGGTCTCTGATGGCAGGAGATGACCCGCCCAATTCAGAGAGAGGTGAAAAACGCC 403
DB 47158 CGGAGATAGTCTCTGATGTCCAGGGGAGATTTCTGCCAAATCAGAGCAATTTAAATGCC 47217
QY 404 TGAAGAAGGT 413
DB 47218 AGAAGGAGGT 47227

RESULT 7
AC002415
LOCUS
DEFINITION
Homo sapiens chromosome X clone bWXD142, HTG 13-JUN-2002
***, 6 unordered pieces.
AC002415
VERSION
AC002415.2 GI:21405635
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 93919)
Chen,B., Brownstein,B.H., States,D.J., Schlesinger,D. and
Mazzarella,R.
Direct Submission
Unpublished (1997)
2 (bases 1 to 93919)
Brownstein,B.H., States,D.J. and Mazzarella,R.
Direct Submission
Submitted (12-AUG-1997) Center for Genetics in Medicine, Box 8232,
Washington University School of Medicine, 4566 Scott Avenue, St.
Louis, MO 63110, USA

COMMENT

On Jun 13, 2002 this sequence version replaced gi:2323256.
Current status of this project is available at:
'http://genome.wustl.edu/cgm/seq_projects.html'
Submitted by:
Ellson Chen,
Advanced Center for Genetic Technology,
Applied Biosystems Division of Perlin Elmer Corp.,
850 Lincoln Center Drive,
Foster City, CA 9404 USA
e-mail: ellson@genseq.apldbio.com

and

Buddy Brownstein,
Center for Genetics in Medicine,
Washington University School of Medicine, Box 8232
4566 Scott Avenue,
St. Louis, MO 63110, USA
e-mail: buddy@genetics.wustl.edu

and

David J. States,
Institute for Biomedical Computing
Washington University in St. Louis
700 South Euclid Ave.
St. Louis, MO 63108 USA
e-mail: states@bc.wustl.edu.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 9560: contig of 9560 bp in length
* 9561 9660: gap of 100 bp
* 9661 25227: contig of 15567 bp in length
* 25228 25327: gap of 100 bp
* 25328 45875: contig of 20548 bp in length
* 45876 45975: gap of 100 bp
* 45976 73699: contig of 27724 bp in length
* 73700 73799: gap of 100 bp
* 73800 82475: contig of 8676 bp in length
* 82476 82575: gap of 100 bp
* 82576 93919: contig of 11344 bp in length.

FEATURES
source

1..93919
/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/clones="bWXD142"

BASE COUNT 25991 a 21189 c 20209 g 26030 t 500 others
ORIGIN

Query Match 12.9%; Score 69.2; DB 1; Length 93919;
Best Local Similarity 70.8%; Pred. No. 9.9e-06;
Matches 92; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 284 AGGCCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCAACCACAGACTGGGTGTGAGTG 343
|||||
Db 59821 AGTCCTTAACCTGGGAAGCTGATCTCCAGAGCTGTCTCAGTCAAGACTGGGGATGAATG 59880
|||||

QY 344 TGAAGATGTCCTGTATGGCAGGAGATGCCGCCCAAAATCCAGAGGAGGTGAAAAACGCC 403
|||||
Db 59881 CGGAGATAGTCTGTATGCCAGGGAAGATTCTGCCAAAATCAGAGCAATTTAAAAATGCC 59940
|||||

QY 404 TGAAGAGAGGT 413
|||||

Db 59941 AGAAGGAGGT 59950

RESULT 8

AL445227

LOCUS

DEFINITION

Homo sapiens chromosome X clone RP13-97115, *** SEQUENCING IN

PROGRESS ***, 6 unordered pieces.

ACCESSION

AL445227

VERSION

AL445227.7 GI:11611395

KEYWORDS

HTG; HTGS_PHASE1; HTGS_CANCELLED.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 114344)

McLay,K.

Direct Submission

Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Dec 9, 2000 this sequence version replaced gi:11493333.

----- Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: bb97115

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 112482 bases at least Q40

Consensus quality: 113245 bases at least Q30

Consensus quality: 113597 bases at least Q20

Insert size: 113844; sum-of-contigs

Insert size: 122396; 9.8% error; agarose-fp

Quality coverage: 5.91x in Q20 bases; sum-of-contigs Quality

coverage: 5.72x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 6 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 13809: contig of 13809 bp in length

* 13810 13909: gap of 100 bp

* 13910 51507: contig of 37598 bp in length

* 51508 51607: gap of 100 bp

```
* 51608 65679: contig of 14072 bp in length
* 65680 65779: gap of 100 bp
* 65780 69064: contig of 3285 bp in length
* 69065 69164: gap of 100 bp
* 69165 101518: contig of 32354 bp in length
* 101519 101618: gap of 100 bp
* 101619 114344: contig of 12726 bp in length.
```

FEATURES

```
source
1..114344
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP13-97115"
/clone_lib="RPC1-13.1"
1..13809
/note="assembly_fragment:00241
fragment_chain:1
clone_end:SP6
vector_side:left"
13910..51507
/note="assembly_fragment:00680
fragment_chain:1"
51608..65679
/note="assembly_fragment:00207
fragment_chain:1"
65780..69064
/note="assembly_fragment:00300"
69165..101518
/note="assembly_fragment:01024"
101619..114344
/note="assembly_fragment:01825
clone_end:T7
vector_side:right"
32658 a 24338 c 24232 g 32616 t 500 others
```

BASE COUNT

```
Query Match 12.9%; Score 69.2; DB 1; Length 114344;
Best Local Similarity 70.8%; Pred. No. 9.9e-06;
Matches 92; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

```
QY 284 AGGCCGAGCCTGAGCTCATAGCAGGACAGGTCACCCACACACTGGGTGAGTG 343
Db 39620 AGTGCCCTAACCTGGAAGCTGATCTCCAGGAGCTGCTCAGTCAAGACTGGGGATGAATG 39679

QY 344 TGAAGATGTCCTGATGGCGAGGAGATGGACCGCCAAATCCAGAGAGGTGAAACGCC 403
Db 39680 CGGAGATAGTCCTGATGTCGAGGGAAGATTCTGCCAAATCAGACGCAATTTAAATGCC 39739

QY 404 TGAAGAGGT 413
Db 39740 AGAAGGAGGT 39749
```

RESULT 9

```
AL445227/c
LOCUS AL445227 114344 bp DNA linear HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome X clone RP13-97115, *** SEQUENCING IN
PROGRESS ***, 6 unordered pieces.
ACCESSION AL445227
VERSION AL445227.7 GI:11611395
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
```

ORGANISM

```
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 114344)
```

REFERENCE

```
1 Mclay,K.
Direct Submission
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
```

COMMENT

```
On Dec 9, 2000 this sequence version replaced gi:11493333.
----- Genome Center
```

```
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bB97115
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 112482 bases at least Q40
Consensus quality: 113245 bases at least Q30
Consensus quality: 113597 bases at least Q20
Insert size: 113844; sum-of-contigs
Insert size: 122396; 9.8% error; agarose-fp
Quality coverage: 5.91x in Q20 bases; sum-of-contigs Quality
coverage: 5.72x in Q20 bases; agarose-fp
-----
```

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
* 1 13809: contig of 13809 bp in length
* 13810 13909: gap of 100 bp
* 13910 51507: contig of 37598 bp in length
* 51508 51607: gap of 100 bp
* 51608 65679: contig of 14072 bp in length
* 65680 65779: gap of 100 bp
* 65780 69064: contig of 3285 bp in length
* 69065 69164: gap of 100 bp
* 69165 101518: contig of 32354 bp in length
* 101519 101618: gap of 100 bp
* 101619 114344: contig of 12726 bp in length.
```

FEATURES

source

```
1..114344
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP13-97115"
/clone_lib="RPC1-13.1"
1..13809
/note="assembly_fragment:00241
fragment_chain:1
clone_end:SP6
vector_side:left"
13910..51507
/note="assembly_fragment:00680
fragment_chain:1"
51608..65679
/note="assembly_fragment:00207
fragment_chain:1"
65780..69064
/note="assembly_fragment:00300"
69165..101518
/note="assembly_fragment:01024"
101619..114344
/note="assembly_fragment:01825
clone_end:T7
vector_side:right"
```

misc_feature

```
/note="assembly_fragment:00241
fragment_chain:1
clone_end:SP6
vector_side:left"
```

misc_feature

```
/note="assembly_fragment:00680
fragment_chain:1"
51608..65679
/note="assembly_fragment:00207
fragment_chain:1"
```

misc_feature

```
/note="assembly_fragment:00300"
69165..101518
/note="assembly_fragment:01024"
101619..114344
/note="assembly_fragment:01825
clone_end:T7
vector_side:right"
```

misc_feature

```
/note="assembly_fragment:00207
fragment_chain:1"
65780..69064
/note="assembly_fragment:00300"
69165..101518
/note="assembly_fragment:01024"
101619..114344
/note="assembly_fragment:01825
clone_end:T7
vector_side:right"
```

misc_feature

```
/note="assembly_fragment:01024"
101619..114344
/note="assembly_fragment:01825
clone_end:T7
vector_side:right"
```

```
BASE COUNT 32658 a 24338 c 24232 g 32616 t 500 others
ORIGIN
```

Query Match

```
Best Local Similarity 12.9%; Score 69.2; DB 1; Length 114344;
Matches 92; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
```

Qy 284 AGGCCGAGCCTGAGCTCATAGCAGGACAGGTCACCCACACACTGGGTGAGTG 343

Db 92320 AGTGCCCTAACCTGGAAGCTGATCTCCAGGAGCTGCTCAGTCAAGACTGGGGGTGAATG 92361


```

* 12396 13579: contig of 1184 bp in length
* 13580 13679: gap of unknown length
* 13680 15201: contig of 1522 bp in length
* 15201 15301: gap of unknown length
* 15301 16390: contig of 1089 bp in length
* 16390 16490: gap of unknown length
* 16490 18667: contig of 2177 bp in length
* 18667 18768: gap of unknown length
* 18768 20869: contig of 2102 bp in length
* 20869 20969: gap of unknown length
* 20969 23568: contig of 2599 bp in length
* 23568 23689: gap of unknown length
* 23689 24252: contig of 584 bp in length
* 24252 28943: gap of unknown length
* 28943 29043: contig of 4591 bp in length
* 29043 37619: gap of unknown length
* 37619 49927: contig of 8576 bp in length
* 49927 50027: contig of 12208 bp in length
* 50027 69100: gap of unknown length
* 69100 69200: contig of 19073 bp in length
* 69200 90260: gap of unknown length
* 90260 90360: contig of 21060 bp in length
* 90360 90361: gap of unknown length
* 90361 141787: contig of 51426 bp in length
* 141787 180859: gap of unknown length
* 180859 180859: contig of 38973 bp in length.

```

FEATURES

source

```

Location/Qualifiers
1. .180859
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-485B17"
1. .1341
/note="assembly_name:Contig19"
1442. .2824
/note="assembly_name:Contig24"
2925. .4091
/note="assembly_name:Contig25"
4192. .5300
/note="assembly_name:Contig26"
5401. .6531
/note="assembly_name:Contig40"
6632. .7946
/note="assembly_name:Contig44"
8047. .9816
/note="assembly_name:Contig45"
9917. .11066
/note="assembly_name:Contig46"
11167. .12295
/note="assembly_name:Contig47"
12396. .13579
/note="assembly_name:Contig48"
13680. .15201
/note="assembly_name:Contig49"
15302. .16390
/note="assembly_name:Contig50"
16491. .18667
/note="assembly_name:Contig52"
18768. .20869
/note="assembly_name:Contig53"
20970. .23568
/note="assembly_name:Contig54
clone end:SP6
vector_side:right"
23669. .24252
/note="assembly_name:Contig56"
24353. .28943
/note="assembly_name:Contig55"
29044. .37619
/note="assembly_name:Contig56"
37720. .49927
/note="assembly_name:Contig57"

```

```

misc_feature 50028. .69100
/note="assembly_name:Contig58"
misc_feature 69201. .90260
/note="assembly_name:Contig59"
misc_feature 90361. .141786
/note="assembly_name:Contig60"
misc_feature 141887. .180859
/note="assembly_name:Contig61"
BASE COUNT 53372 a 37170 c 37454 g 50639 t 2224 others
ORIGIN

```

```

Query Match 12.6%; Score 67.6; DB 1; Length 180859;
Best Local Similarity 70.0%; Pred. No. 2.2e-05;
Matches 91; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 284 AGGGCCGAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTG 343
Db 93228 AGTGCCTGACCTGGAAGCTGATCTCCAGGAGCTGCTCAGTCAAGAAGCTGGGATGAATG 93287
QY 344 TGAAGATGCTCTGATGGGAGGAGATGGACCCGCAATCCAGAGGAGGTGAACAGCC 403
Db 93288 CGAGATGGTCTGATGTCAGGGGAAAGATTCTGACAAAGTCAGAGCAATTAAATGCC 93347
QY 404 TGAAGAAAGGT 413
Db 93348 AGAAGGAGGT 93357

```

RESULT 12

```

AL772246 171489 bp DNA linear HTG 13-AUG-2002
LOCUS Homo sapiens chromosome X clone RP13-3418, *** SEQUENCING IN
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AL772246
VERSION AL772246.4 GI:22265477
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Whitehead,S.
Direct Submission
Submitted (10-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 15, 2002 this sequence version replaced gi:22002760.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BB3418
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 171482 bases at least Q40
Consensus quality: 171488 bases at least Q30
Consensus quality: 171489 bases at least Q20
Insert size: 171489; sum-of-contigs
Insert size: 186799; 2.2% error; agarose-fp
Quality coverage: 12.50x in Q20 bases; sum-of-contigs Quality
coverage: 11.91x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

FEATURES

source

```

1. .171489
/organism="Homo sapiens"
/db_xref="taxon:9606"

```


* 13988 14087: gap of 100 bp
* 14088 14769: contig of 682 bp in length
* 14770 14869: gap of 100 bp
* 14870 15548: contig of 679 bp in length
* 15549 15648: gap of 100 bp
* 15649 16324: contig of 676 bp in length
* 16325 16424: gap of 100 bp
* 16425 17108: contig of 684 bp in length
* 17109 17208: gap of 100 bp
* 17209 17892: contig of 684 bp in length
* 17893 17992: gap of 100 bp
* 17993 18672: contig of 680 bp in length
* 18673 18772: gap of 100 bp
* 18773 19475: contig of 703 bp in length
* 19476 19575: gap of 100 bp
* 19576 20270: contig of 695 bp in length
* 20271 20370: gap of 100 bp
* 20371 21027: contig of 657 bp in length
* 21028 21127: gap of 100 bp
* 21128 21804: contig of 677 bp in length
* 21805 21944: gap of 100 bp
* 21905 22597: contig of 693 bp in length
* 22598 22697: gap of 100 bp
* 22698 23391: contig of 694 bp in length
* 23392 23491: gap of 100 bp
* 23492 24200: contig of 709 bp in length
* 24201 24300: gap of 100 bp
* 24301 24983: contig of 683 bp in length
* 24984 25083: gap of 100 bp
* 25084 25777: contig of 694 bp in length
* 25778 25877: gap of 100 bp
* 25878 26560: contig of 683 bp in length
* 26561 26660: gap of 100 bp
* 26661 27355: contig of 699 bp in length
* 27360 27459: gap of 100 bp
* 27460 28128: contig of 669 bp in length
* 28129 28228: gap of 100 bp
* 28229 28896: contig of 668 bp in length
* 28897 28996: gap of 100 bp
* 28997 29663: contig of 667 bp in length
* 29664 29763: gap of 100 bp
* 29764 30454: contig of 691 bp in length
* 30455 30554: gap of 100 bp
* 30555 31232: contig of 678 bp in length
* 31233 31332: gap of 100 bp
* 31333 32028: contig of 696 bp in length
* 32029 32128: gap of 100 bp
* 32129 32816: contig of 688 bp in length
* 32817 32916: gap of 100 bp
* 32917 33595: contig of 679 bp in length
* 33596 33695: gap of 100 bp
* 33696 34380: contig of 685 bp in length
* 34381 34480: gap of 100 bp
* 34481 35189: contig of 709 bp in length
* 35190 35289: gap of 100 bp
* 35290 35957: contig of 668 bp in length
* 35958 36057: gap of 100 bp
* 36058 36734: contig of 677 bp in length
* 36735 36834: gap of 100 bp
* 36835 37502: contig of 668 bp in length
* 37503 37602: gap of 100 bp
* 37603 38286: contig of 684 bp in length
* 38287 38386: gap of 100 bp
* 38387 39060: contig of 674 bp in length
* 39061 39160: gap of 100 bp
* 39161 39850: contig of 690 bp in length
* 39851 39950: gap of 100 bp
* 39951 40656: contig of 706 bp in length
* 40657 40756: gap of 100 bp
* 40757 41452: contig of 696 bp in length
* 41453 41552: gap of 100 bp
* 41553 42260: contig of 708 bp in length
* 42261 42360: gap of 100 bp

* 42361 43028: contig of 668 bp in length
* 43029 43128: gap of 100 bp
* 43129 43800: contig of 672 bp in length
* 43801 43900: gap of 100 bp
* 43901 44566: contig of 666 bp in length
* 44567 44666: gap of 100 bp
* 44667 45354: contig of 688 bp in length
* 45355 45454: gap of 100 bp
* 45455 46137: contig of 683 bp in length
* 46138 46237: gap of 100 bp
* 46238 46917: contig of 680 bp in length
* 46918 47017: gap of 100 bp
* 47018 47676: contig of 659 bp in length
* 47677 47776: gap of 100 bp
* 47777 48461: contig of 685 bp in length
* 48462 48561: gap of 100 bp
* 48562 49248: contig of 687 bp in length
* 49249 49348: gap of 100 bp
* 49349 50024: contig of 676 bp in length
* 50025 50124: gap of 100 bp
* 50125 50802: contig of 678 bp in length
* 50803 50902: gap of 100 bp
* 50903 51607: contig of 705 bp in length
* 51608 51707: gap of 100 bp
* 51708 52385: contig of 678 bp in length
* 52386 52485: gap of 100 bp
* 52486 53131: contig of 646 bp in length
* 53132 53231: gap of 100 bp
* 53232 53905: contig of 674 bp in length

Query Match 11.6%; Score 62.6; DB 1; Length 64891;
Best Local Similarity 66.9%; Pred. No. 0.00029;
Matches 89; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 281 TCAGGGCCGAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGA 340
Db 32401 TTAAGTGTCTACCTGGAGCTGATCTCCAGGATTTGTCTCAGTCAAGACTGGGAATGA 32342
QY 341 GTGTGAAGATGGTCTCTGATGGCGAGAGATGGACCCGCCAAATCCAGAGAGGTGGAATAC 400
Db 32341 ATGCAGAGATGATCTCTGATGTCAAGGGGAAGATTCTGCCAAACTAGAGCATTTTAAAT 32282
QY 401 GCCTGAAGAAGGT 413
Db 32281 GCCAGACGAGGT 32269

RESULT 15

AC002415/c

LOCUS

DEFINITION

AC002415

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AC002415 Homo sapiens chromosome X clone bx142, linear HTG 13-JUN-2002
***, 6 unordered pieces.
AC002415 AC002415.2 GI:21405635
HTGS_PHASE1.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 93919)
Chen, E., Brownstein, B.H., States, D.J., Schlessinger, D. and
Mazzarella, R.
Direct Submission
Unpublished (1997)
2 (bases 1 to 93919)
Brownstein, B.H., States, D.J. and Mazzarella, R.
Submitted (12-AUG-1997) Center for Genetics in Medicine, Box 8232,
Washington University School of Medicine, 4566 Scott Avenue, St.
Louis, MO 63110, USA
On Jun 13, 2002 this sequence version replaced gi:2323256.
Current status of this project is available at:
'http://genome.wustl.edu/cgm/seq_projects.html'

Search completed: April 9, 2003, 16:53:24
Job time : 1804 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2003, 15:17:04 ; Search time 1530 Seconds
(without alignments)
5694.884 Million cell updates/sec

Title: US-09-782-745-14
Perfect score: 538
Sequence: 1 ACCCGAGGAGCTGTGAGGC.....CTGCAAGAAAAA 538

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST.*
1: em_estba.*
2: em_esthum.*
3: em_estmu.*
4: em_estm.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_estl.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	524.6	97.5	569	12	BG700165
2	513	95.4	537	14	BM832793
3	492.6	91.6	813	12	BG773070
4	492	91.4	527	13	B1826605
5	492	91.4	851	12	BG250953
6	490.6	91.2	572	12	BG120838

7	476.8	88.6	509	13	B1868671
8	463.4	86.1	623	13	BM172302
9	452.4	84.1	489	9	AA447559
10	431.4	80.2	457	10	AW510753
11	424.6	78.9	447	9	A1381509
12	424.4	78.9	464	9	AA738037
13	424	78.8	581	12	BG700088
14	413.8	76.9	418	10	AW016546
15	413.4	76.8	480	12	BG120336
16	412.8	76.7	426	9	AA868226
17	410.8	76.4	450	9	AA760996
18	401.2	74.6	455	9	A1187350
19	370.4	68.8	795	13	B1S60935
20	365.4	67.9	412	10	AW102587
21	360.4	67.0	445	11	BC005363
22	359.4	66.8	419	9	AA448542
23	357	66.4	383	14	BM836228
24	346.6	64.4	839	12	BG527731
25	329.8	61.3	384	9	AA913206
26	318.8	59.3	397	12	BG206349
27	309.8	57.6	1069	12	BE777698
28	296.4	55.1	398	9	AA918604
29	290.6	54.0	320	14	BM836021
30	250.4	46.5	256	12	BG181480
31	233.4	43.4	258	12	BG186708
32	228.2	42.1	253	12	BG184057
33	226.6	42.1	261	12	BG199060
34	224.8	41.8	245	12	BG212621
35	219.8	40.9	292	12	BG212622
36	214.8	39.9	245	12	BG208433
37	213.6	39.7	849	12	BG436305
38	213.4	39.7	524	11	AF318372
39	207.6	38.6	507	9	AA972716
40	207	38.5	256	12	BG216461
41	207	38.5	547	14	AA0147
42	204.8	38.1	224	9	AA738394
43	202.4	37.6	382	12	BF869799
44	200.6	37.3	499	12	BG354572
45	193.8	36.0	197	9	A1968311

ALIGNMENTS

RESULT 1
BG700165 569 bp mRNA linear EST 07-MAY-2001
602679622F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4812437 5',
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BG700165 569 bp mRNA linear EST 07-MAY-2001
602679622F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4812437 5',
mRNA sequence.
BG700165
BG700165.1 GI:13969233
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 569)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Incyte Genomics, Inc.
Found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10704 row: 1 column: 06
High quality sequence stop: 566.
Location/Qualifiers

FEATURES

```

source
1. .569
/organism="Homo sapiens"
/db xref="taxon:9606"
/clonelib="IMAGE:4812437"
/clonelib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab host="DH10B"
/notes="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.5 kb and
normalized to R0F 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NIHRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 166 a 125 c 165 g 113 t
ORIGIN
Query Match 97.5%; Score 524.6; DB 12; Length 569;
Best Local Similarity 99.2%; Pred. No. 2.5e-118;
Matches 527; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 8 GGAGCTGTGAGGAGTGTGTGGTCTCTGCGTCCGACCTTTTCTCTACTGAGA 67
Db 25 GGAGCTGTGAGGAGTGTGTGGTCTCTGCGTCCGACCTTTTCTCTACTGAGA 84
QY 68 TTCAATCTGTGAAATATGATGTCGGGAGGAGATCGACCTATCGGCTAGACCAAGACG 127
Db 85 TTCAATCTGTGAAATATGATGTCGGGAGGAGATCGACCTATCGGCTAGACCAAGACG 144
QY 128 CTACGTAGAGCTCTCTGAATGATTTGGCCCTATGCGGCCGAGCAGTTCAGTGATGAAGT 187
Db 145 CTACGTAGAGCTCTCTGAATGATTTGGCCCTATGCGGCCGAGCAGTTCAGTGATGAAGT 204
QY 188 GGAACAGACACCTGAAAGAGGGGAAACCGAACTCAACGTGAGGATCTCTGCGAGTGC 247
Db 205 GGAACAGACACCTGAAAGAGGGGAAACCGAACTCAACGTGAGGATCTCTGCGAGTGC 264
QY 248 TCAGGAGGAGGAGATGAGGAGGATCTGCAAGGTCAAGGGCCGAGCCTGAAGCTCATAG 307
Db 265 TCAGGAGGAGGAGATGAGGAGGATCTGCAAGGTCAAGGGCCGAGCCTGAAGCTCATAG 324
QY 308 CAGGAGCAGGCTACCCACAGATGGGTGTGAGTGTGAAGATGCTGCTGATGGCAGGA 367
Db 325 CAGGAGCAGGCTACCCACAGATGGGTGTGAGTGTGAAGATGCTGCTGATGGCAGGA 384
QY 368 GATGGACCCGCCAAATCCAGAGGAGGTGAAACCGCTCAAGAGGTGAAAAGCAATCACA 427
Db 385 GATGGACCCGCCAAATCCAGAGGAGGTGAAACCGCTCAAGAGGTGAAAAGCAATCACA 444
QY 428 GTGTTAAAGAGACACGTTGAAATGATGCAAGGTGCTTCTATGTTGAAATTTGTTTAT 487
Db 445 GTGTTAAAGAGACACGTTGAAATGATGCAAGGTGCTTCTATGTTGAAATTTGTTTAT 504
QY 488 TAAATTTCTCCCAATAAGCTTTACAGCTTCTGCAAGAGAAAAA 538
Db 505 TAAATTTCTCCCAATAAGCTTTACAGCTTCTGCAAGAGAAAAA 555

RESULT 2
BM832793
LOCUS
DEFINITION
X-EST0107334 SSSNU484s1 Homo sapiens cDNA clone SSSNU484s1-8-F10
5', mRNA sequence.
ACCESSION
BM832793
VERSION
BM832793.1 GI:19189202
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 537)

```

```

AUTHORS
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 8 row: F column: 10
High quality sequence stop: 537.
Location/Qualifiers
1. .537
/organism="Homo sapiens"
/db xref="taxon:9606"
/clonelib="SSSNU484s1-8-F10"
/clonelib="SSSNU484s1"
/sex="M"
/tissue_type="Stomach"
/cell_line="SNU-484"
/lab host="Top10F"
/notes="Organ: Stomach; Vector: pTZ18RP1; Site 1: EcoRI;
Site 2: NotI; The poly (A)+ RNA was decapped with tobacco
acid pyrophosphatase (TAP) and ligated with DNA-RNA linker
including EcoRI site by treatment of T4 RNA ligase. The
first strand cDNA was synthesized from oligo dT-selected
mRNA by priming with dT-tailed vector. The dT-tailed
vector was adjusted to have about 60nt. The cDNA vector
was circularized with E. coli DNA ligase after digestion
of EcoRI which site is also included in vector. An RNA
strand converted to a DNA strand by Okayama-Berg method.
The obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
After analyzing and sequencing about 2,000 ~ 3,000
colonies in original cDNA library, the abundant cDNAs were
selected and amplified by PCR reaction using vector region
primer including T7 promoter as 5' primer and N(gt)14 as
3' primer. The PCR products were used as template for
synthesis of biotinylated single stranded RNA by in vitro
transcription reaction. The synthesized RNA probes were
hybridized with antisense single stranded cDNAs prepared
from original library and incubated with avidin-gel.
After removing DNA-RNA hybrids by centrifuge, the
subtracted cDNA libraries were constructed by
transformation of the remaining DNA into competent cells E.
coli Top10F with electroporation method."
BASE COUNT 156 a 114 c 156 g 111 t
ORIGIN
Query Match 95.4%; Score 513; DB 14; Length 537;
Best Local Similarity 98.5%; Pred. No. 1.8e-115;
Matches 529; Conservative 0; Mismatches 5; Indels 3; Gaps 1;
QY 3 GCCAGGAGCTGTGAGGAGTGTGTGGTTCCTGCGCTCGGACTCTTTTCTCTAC 62
Db 1 GCCAGGAGCTGTGAGGAGTGTGTGGTTCCTGCGCTCGGACTCTTTTCTCTAC 60
QY 63 TGAGATTTCATCTGTGTAATATGAGTTGGCGAGGAAGATCGACCC 119
Db 61 TGAGATTTCATCTGTGTAATATGAGTTGGCGAGGAAGATCGACCC 120
QY 120 CCAAGAGCTGTGAGGAGCTCTGAAATGATTTGGGCTATGCGGCCGAGGATTCAGT 179
Db 121 CCAAGGCGCTATGTACAGCTCTCTGAAGTGTGAGGCTATGCGGCCGAGGATTCAGT 180
QY 180 GATGAAGTGAACCCAGCAACACCTGAAAGAGGGGAACAGCAACTCAACGTCAGGATCCT 239
Db 181 GATGAAGTGAACCCAGCAACACCTGAAAGAGGGGAACAGCAACTCAACGTCAGGATCCT 240

```

```

TITLE
JOURNAL
COMMENT

```

```

QY 240 GCAGCTGCTCAGGAGGAGGATGAGGAGCATCTGCAGGTCAAAGGCCGAGCCTGAA 299
Db 241 GCAGCTGCTCAGGAGGAGGATGAGGAGCATCTGCAGGTCAAAGGCCGAGCCTGAA 300
QY 300 GCTCATAGCCAGGAAACAGGGTCAACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTTGAT 359
Db 301 GCTCATAGCCAGGAAACAGGGTCAACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTTGAT 360
QY 360 GGGCAGGAGATGAGCCCGCAATCCAGGAGGTGAAGCCCTGAAGAGGTGAAG 419
Db 361 GGGCAGGAGATGAGCCCGCAATCCAGGAGGTGAAGCCCTGAAGAGGTGAAG 420
QY 420 CAATCACAGCTGTTAAAGAACACAGCTTGAATGATCAGGCTGCTCTCTATGTTGGAAT 479
Db 421 CAATCACAGCTGTTAAAGAACACAGCTTGAATGATCAGGCTGCTCTCTATGTTGGAAT 480
QY 480 TTGTTTCATTAAATTCCTCCCAATAAGCTTTACAGCCTTTGCAAGAAAAA 536
Db 481 TTGTTTCATTAAATTCCTCCCAATAAGCTTTACAGCCTTTGCAAGAAAAA 537

RESULT 3
BG773070 813 bp mRNA linear EST 15-MAY-2001
LOCUS 602721370F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4838176 5',
DEFINITION mRNA sequence.
ACCESSION BG773070.1 GI:14083723
VERSION BG773070.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 813)
NIH-MGC http://imgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rcgaps@rmail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLMI0771 row: 1 column: 17
High quality sequence stop: 778.
Location/Qualifiers
1..813
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4838176"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcag
); Oligo-dr primed using primer 5'-TTTCTTTTCTTTTCTTN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 215 a 163 c 209 g 226 t
ORIGIN
Query Match 91.6%; Score 492.6; DB 12; Length 813;
Best Local Similarity 96.8%; Pred. No. 1.8e-110;
Matches 514; Conservative 0; Mismatches 14; Indels 3; Gaps 1;

QY 1 ACGCCAGGAGCTGTGAGSCAGTGTGTGGTCTCTCGCGTCTCTTTTCTCT 60

```

```

Db 10 ACGCCAGGAGCTGTGAGSCAGTGTGTGGTCTCTCGCGTCTCTTTTCTCT 69
QY 61 ACTGAGATTTCATCTGTGTAATATGATGTTGGCCGAGGAAGATCGACC--TATCGGCCTA 117
Db 70 ACTGAGATTTCATCTGTGTAATATGATGTTGGCCGAGGAAGATCGACCCTATTTATTTGGCCTA 129
QY 118 GACCAAGACGCTACGTAGAGCCTCTCTGAAATGATTGGGCCCTATGCGGCCGAGCAGTTCA 177
Db 130 GACCAAGGCGCTATGTACAGCCTCTCTGAAATGATTGGGCCCTATGCGGCCGAGCAGTTCA 189
QY 178 GTGATGAAGTGAACCGAGCAACCTCAAGNAGGGGACCCAGCACTCAACCTCAGGATC 237
Db 190 GTGATGAAGTGAACCGAGCAACCTCAAGNAGGGGACCCAGCACTCAACCTCAGGATC 249
QY 238 CTCAGCTGCTCAGGAGGAGGATGAGGAGGATCTGCAGGTCAAGGCCGAGCCTG 297
Db 250 CTCAGCTGCTCAGGAGGAGGATGAGGAGGATCTGCAGGTCAAGGCCGAGCCTG 309
QY 298 AAGCTCATAGCCAGGAAACAGGGTCAACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTG 357
Db 310 AAGCTCATAGCCAGGAAACAGGGTCAACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTG 369
QY 358 ATGGCAGGAGATGACCCGCCAATCCAGAGGAGGTGAAGCCCTGAGAGGTGAAA 417
Db 370 ATGGCAGGAGATGACCCGCCAATCCAGAGGAGGTGAAGCCCTGAGAGGTGAAA 429
QY 418 AGCAATCACAGTGTAAAGAAAGACACCGTTGAAATGATGTCAGGCTCTCTCTATGTTGGAA 477
Db 430 GGCATCACAGTGTAAAGAAAGACATCTGTAATGTTGTCAGGCTCTCTCTATGTTGGAA 489
QY 478 ATTGTTTCATTAAATTCCTCCCAATAAGCTTTACAGCCTTTGCAAGAAA 528
Db 490 AATTCTTCATTAAAGTCTCCCAATAAGCTTTACAGCCTTTCTGCAAGAAA 540

RESULT 4
BG773070 527 bp mRNA linear EST 04-OCT-2001
LOCUS 603077056F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5168892 5',
DEFINITION mRNA sequence.
ACCESSION BG773070.1 GI:15938155
VERSION BG773070.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 527)
NIH-MGC http://imgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rcgaps@rmail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLMI1419 row: p column: 13
High quality sequence stop: 519.
Location/Qualifiers
1..527
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5168892"
/clone_lib="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: EcoRV (destroyed); RNA source normal medulla from
source
FEATURES

```

anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."

BASE COUNT	154 a	111 c	151 g	111 t
ORIGIN				
Query Match	91.4%; Score 492; DB 13; Length 527;			
Best Local Similarity	98.3%; Pred. No. 2.5e-110;			
Matches	519; Conservative 0; Mismatches 5; Indels 4; Gaps 2;			
QY	12	CTGTGAGGCGAGTCTGTGGTTCCTGCGTCCGGACTCTTTTCCCTACTGAGATTCA	71	
Db	1	CTGTGAGGCGAGTCTGTGGTTCCTGCGTCCGGACTCTTTTCCCTACTGAGATTCA	60	
QY	72	TCTGTGAAATATAGTGTGGGAGGAATCGAAC---TATCGGCCTAGACCAAGAGCG	128	
Db	61	TCTGTGAAATATAGTGTGGGAGGAATCGAACCTATTATTGGCCTAGACCAAGGCG	120	
QY	129	TACGTAGAGCTCTCGAATAGTGGGCTATGCGGCCCGGACAGTTTCAGTGATGAAGTG	188	
Db	121	TATGTACAGCTCTCGAATAGTGGGCTATGCGGCCCGGACAGTTTCAGTGATGAAGTG	180	
QY	189	GAACGAGCAACACCTGGAAGAGGGAACGAGCAACTCAACGTCAGGATCTCTGCAAGTCT	248	
Db	181	GAACGAGCAACACCTGGAAGAGGGAACGAGCAACTCAACGTCAGGATCTCTGCAAGTCT	240	
QY	249	CAGGAGGAGGAGTATGAGGAGCATCTGCAGGTCAAGGGCCGAGCCTGAAAGTCTATAGC	308	
Db	241	CAGGAGGAGGAGTATGAGGAGCATCTGCAGGTCAAGGGCCGAGCCTGAAAGTCTATAGC	300	
QY	309	CAGGAACAGGCTCACCCACAGACTGGTGTGAGTGTGAAGATGCTCTGATGGCAGGAG	368	
Db	301	CAGGAACAGGCTCACCCACAGACTGGTGTGAGTGTGAAGATGCTCTGATGGCAGGAG	360	
QY	369	ATGACCCGCCAATCCAGAGGAGTGAACCGCTTGAAAGAGGTGAAAGCAATCACAG	428	
Db	361	ATGACCCGCCAATCCAGAGGAGTGAACCGCTTGAAAGAGGTGAAAGCAATCACAG	420	
QY	429	TGTTAAAGAGACACCTTGAATGATGAGGCTGCTCTATGTTGAAATTTGTTCAAT	488	
Db	421	TGTTAAAGAGGACACCTTGAATGATGAGGCTGCTCTATGTTGAAATTTGTTCAAT	480	
QY	489	AAATTTCTCCCAATAAAGCTTTTACAGCTTCTGCAAGAGAAAAA	536	
Db	481	AAATTTCTCCCAATAAAGCTTTTACAG-CTTCTGCAAGAGAAAAA	527	
RESULT 5				
LOCUS	BG250953			
DEFINITION	602363801F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4471969 5',			
ACCESSION	BG250953			
VERSION	BG250953.1			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
JOURNAL	1 (bases 1 to 851)			
COMMENT	NIH-MGC http://mgi.nci.nih.gov/.			
	National Institutes of Health, Mammalian Gene Collection (MGC)			
	Unpublished (1999)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgrabs-r@mail.nih.gov			
	Tissue Procurement: ATCC			
	cDNA Library Preparation: Life Technologies, Inc.			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: Incyte Genomics, Inc.			

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAMI0292 row: b column: 02
High quality sequence stop: 531.
Location/Qualifiers
1..851
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4471969"
/clone_lib="NIH_MGC_90"
/tissue_type="adenoCarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: Noti; Site 2: Sali; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT	224 a	205 c	239 g	183 t
ORIGIN				
Query Match	91.4%; Score 492; DB 12; Length 851;			
Best Local Similarity	96.4%; Pred. No. 2.5e-110;			
Matches	515; Conservative 0; Mismatches 15; Indels 4; Gaps 1;			
QY	8	GGAGCTGTGAGGCGAGTCTGTGGTTCCTGCGCTCCGACTCTTTTCTCTACTGAGA	67	
Db	11	GGAGCTGTGAGGCGAGTCTGTGGTTCCTGCGCTCCGACTCTTTTCTCTACTGAGA	70	
QY	68	TTTCATCTGTGAAATATAGTGTGGCGAGGAAGATCGACCTA---TCGGCCTAGACCAA	123	
Db	71	TTTCATCTGTGAAATATAGTGTGGCGAGGAAGATCGACCTATTATTGGCTAGACCAA	130	
QY	124	GAGCTTACGTAGAGCTCTGAAATGATTGGSCCTATGCGGCCGAGCAAGTTCAAGTGA	183	
Db	131	GCGCTATGTACAGCTCTCTGAAAGTATTGGSCCTATGCGGCCGAGCAAGTTCAAGTGA	190	
QY	184	AAGTGAACACAGCAACACCTGAAAGAGGGGAACAGCAACTCAACCTCAGGCTCTCGAG	243	
Db	191	AAGTGAACACAGCAACACCTGAAAGAGGGGAACAGCAACTCAACCTCAGGCTCTCGAG	250	
QY	244	CTGCTCAGGAGGAGGAGGATGAGGAGCATCTGCAAGGTCAAGGCCGAGCCTGAAGCTC	303	
Db	251	CTGCTCAGGAGGAGGAGGATGAGGAGCATCTGCAAGGTCAAGGCCGAGCCTGAAGCTC	310	
QY	304	ATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGTGTCTCTGATGGC	363	
Db	311	ATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGTGTCTCTGATGGC	370	
QY	364	AGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAACCGCTTCAAGAGGTGAAAGCAAT	423	
Db	371	AGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAACCGCTTCAAGAGGTGAAAGCAAT	430	
QY	424	CACAGTGTAAAGAGAGACAGTTGAAATGATGAGGCTGTCTCTATGTTGAAATTTGT	483	
Db	431	CACAGTGTAAAGAGAGACAGTTGAAATGATGAGGCTGTCTCTATGTTGAAATTTGT	490	
QY	484	TCATTAAATTTCTCCCAATAAAGCTTTACAGCTTCTGCAAGAGAAAAA	537	
Db	491	TCATTAAATTTCTCCCAATAAAGCTTTACAGCTTCTGCAAGAGAAAAA	544	

RESULT 6				
LOCUS	BG120838			
DEFINITION	602351010F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:445882 5',			
ACCESSION	BG120838			
VERSION	BG120838.1			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 572)
NIH-MGC <http://mgs.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10224 row: c column: 03
High quality sequence start: 5
High quality sequence stop: 572.
Location/Qualifiers
1. .572
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4445882"
/clone_lib="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 167 a 124 c 166 g 115 t
ORIGIN

Query Match 91.2%; Score 490.6; DB 12; Length 572;
Best Local Similarity 96.7%; Pred. No. 5.5e-110;
Matches 523; Conservative 0; Mismatches 14; Indels 4; Gaps 2;

QY 1 ACGCAGGAGCTGTGAGGCACTGTGTGTTCTCTGCGTCCGCGACTCTTTTCTCT 60
DB 28 ACGCAGGAGCTGTGAGGCACTGTGTGTTCTCTGCGTCCGCGACTCTTTTCTCT 87

QY 61 ACTGAGATTCACTGTGTGAATATGAGTTGGCGAGGAAGATGACAC---TATCGGCCTA 117
DB 88 ACTGAGATTCACTGTGTGAATATGAGTTGGCGAGGAAGATGACACTATTATTGGCCTA 147

QY 118 GACCAAGACGCTACGTAGAGCTCTGAAATGATTGGGCTATGCGGCCCGAGCAGTTCA 177
DB 148 GACCAAGGCGCTATGTACAGCCTCTCTGAAGTGAATGGGCTATGCGGCCCGAGCAGTTCA 207

QY 178 GTGATGAAGTGGACCAACACCTGAAGAGGGGGAACCAAGCAACTCAACGTCAGGATC 237
DB 208 GTGATGAAGTGGACCAACACCTGAAGAGGGGGAACCAAGCAACTCAACGTCAGGATC 267

QY 238 CTGACGCTGCTCAGAGGAGAGGATGAGGAGCATCTGCAAGTCAAGGCCCGAAGCCTG 297
DB 268 CTGACGCTGCTCAGAGGAGAGGATGAGGAGCATCTGCAAGTCAAGGCCCGAAGCCTG 327

QY 298 AAGCTCATAGCCAGCAAGGCTCACCCACAGACTCGGTGTGAGTGAAGATGCTCTG 357
DB 328 AAGCTCATAGCCAGCAAGGCTCACCCACAGACTCGGTGTGAGTGAAGATGCTCTG 387

QY 358 ATGGCAGAGATGGAACCGCCAAATCCAGAGAGGAGTGAAGACCGCTGAAGAGAGTGA 417
DB 388 ATGGCAGAGATGGAACCGCCCAATCCAGAGAGGAGTGAAGACCGCTGAAGAGTGA 447

QY 418 AGCAATCAGGTTTAAAGAGACACGTTGAAATGATGAGGCTGCTCTTATTTGGAA 477
DB 448 AGCAATCAGGTTTAAAGAGAGGACGTTGAAATGATGAGGCTG-TCTTATGTCGAA 506

QY 478 ATTTGTTTCAATTAATTTCTCCCAATAAGCTTTACAGCCTTCTGCAAGAAAAA 537
DB 507 ATTTGTTTCAATTAATTTCTCCCAATAAGCTTTTACATGCTTCTGAAAAA 566

QY 538 A 538
DB 567 A 567

RESULT 7
BI868671
LOCUS
DEFINITION
BI868671 509 bp mRNA linear EST 11-OCT-2001
603392594F1 NIH_MGC_90 Homo sapiens CDNA clone IMAGE:5402663 5',
mRNA sequence.
BI868671
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 509)
NIH-MGC <http://mgs.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM12026 row: 1 column: 24
High quality sequence stop: 509.
Location/Qualifiers
1. .509
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5402663"
/clone_lib="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 142 a 107 c 154 g 106 t
ORIGIN

Query Match 88.6%; Score 476.8; DB 13; Length 509;
Best Local Similarity 98.0%; Pred. No. 1.3e-106;
Matches 494; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

QY 1 ACGCCAGGAGCTGTGAGGCACTGTGTGTTCTCTGCGTCCGCGACTCTTTTCTCT 60
DB 6 ACGCTAGGAGCTGTGAGGCACTGTGTGTTCTCTGCGTCCGCGACTCTTTTCTCT 65

QY 61 ACTGAGATTCACTGTGTGAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTA 117
DB 66 ACTGAGATTCACTGTGTGAATATGAGTTGGCGAGGAAGATCGACCATTATTGGCCTA 125

QY 118 GACCAAGACGCTACGTAGAGCTCTGAAATGATTGGCCTATGCGGCCCGAGCAGTTCA 177
DB 126 GACCAAGGCGCTATGTACAGCCTCTCTGAAATGATTGGCCTATGCGGCCCGAGCAGTTCA 185

QY 178 GTGATGAAGTGGAAACCAAGCAACACCTGAAGAGGGGGAACCAAGCAACTCAACGTCAGGATC 237
DB 186 GTGATGAAGTGGAAACCAAGCAACACCTGAAGAGGGGGAACCAAGCAACTCAACGTCAGGATC 245

QY 238 CTGACGCTGCTCAGGAGGAGAGGATGAGGAGCATCTGCAAGGTCAGAGGCGCAAGCCTG 297
DB 246 CTGACGCTGCTCAGGAGGAGAGGATGAGGAGCATCTGCAAGGTCAGAGGCGTCAGAGCCTG 305

QY 298 AAGCTCATAGCCAGGAACAGGTCACCCACAGACTGGGTGTGAGTGTGAGATGCTCTG 357

```

|||||
Db 306 AAGCTCATAGCCAGGAACAGGGTCAACCCAGAGCTGGGTGTGAGTGTGAAGATGCTCTCTG 365
QY 358 ATGGCAGAGATGACCCGCCAAATCCAGAGAGGTGAAAACGCTGAAGAGGTGAAA 417
Db 366 ATGGCAGAGATGACCCGCCAAATCCAGAGAGGTGAAAACGCTGAAGAGGTGAAA 425
QY 418 AGCAATCAGCTGTTAAAAAGACACAGCTTGAATGATGAGGCTGCTCTATGTTGAA 477
Db 426 AGCAATCAGCTGTTAAAAAGAGGACGCTTGAATGATGAGGCTGCTCTATGTTGAA 485
QY 478 ATTGTTTCATTAAATCTCCCA 501
Db 486 ATTGTTTCATTAAATCTCCCA 509

RESULT 8
LOCUS BM172302 623 bp mRNA linear EST 04-DEC-2001
DEFINITION imagec_4_2001/sm188bdf41.x1 NIH_MGC_97 Homo sapiens cDNA clone
IMAGE:4838176 5', mRNA sequence.
ACCESSION BM172302
VERSION BM172302.1 GI:17311865
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Kale,P.J., Harsch,T.J., Folta,P.A., Nelson,D.O., Sanders,C.G. and
Prange,C.K.
JOURNAL The I.M.A.G.E. Consortium quality control effort: clone
COMMENT resequencing for verification
Other ESTs: BG773070
Contact: Prange CK
The I.M.A.G.E. Consortium
Lawrence Livermore National Laboratory
Livermore, CA, USA
Email: help@image.llnl.gov
This read has been verified (found to hit its original self in the
correct orientation), as part of the I.M.A.G.E. Consortium quality
control effort. High quality sequence is defined as having 100 or
more base pairs with a phred quality value of 20 or greater, where
a sliding window of 4 base pairs with a phred quality value of 15
or greater marks the beginning and end of the sequence. For
information on obtaining this clone, please contact
info@image.llnl.gov.
Plate: L1AM10771 row: 1 column: 17
Seq primer: -21m13
High quality sequence stop: 623.
FEATURES
Location/Qualifiers
1..623
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4838176"
/lab_host="NIH_MGC_97"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 166 a 125 c 173 g 159 t
ORIGIN

Query Match 86.1%; Score 463.4; DB 13; Length 623;
Best Local Similarity 94.5%; Pred. No. 2.6e-103;
Matches 502; Conservative 0; Mismatches 26; Indels 3; Gaps 2;

```

```

QY 1 ACGCCAGGAGCTGTGAGCAGTGTCTGTGTCTCTGCCGTCGGACTCTTTTCTCTCT 60
Db 10 ACGCCAGGAGCTGTGAGCAGTGTCTGTGTCTCTGCCGTCGGACTCTTTTCTCTCT 69
QY 61 ACTCAGATTTCATCTGTGTGAATATGAGTTGGGAGGAAGATCGACCTAT--CGSCCTAG 118
Db 70 ACTCAGATTTCATCTGTGTGAATATGAGTTGGGAGGAAGATCGACCTATTTATTGGCCTA 129
QY 119 ACCAAGACGCTACGTAGAGCCTCTCTGAAATGATTGGGCCCTATGCGGCCGAGCAGTTT 178
Db 130 TATCATGCTCTATGTACAGCCTCTCTGAAATGATTGGGCCCTATGCGGCCGAGCAGTTT 189
QY 179 TGATGAAGTGAACACGCAACACCTGAAGAGGGGAAACCACTCAACTCAAGCTCAGATCC 238
Db 190 TGATGAAGTGAACACGCAACACCTGAAGAGGGGAAACCACTCAAGCTCAGATCC 249
QY 239 TGCAGCTGTCTCAGGAGGAGGAGGATGAGGAGCATCTGCAGGTCAAGGGCCGAGGCTTGA 298
Db 250 TGCAGCTGTCTCAGGAGGAGGAGGATGAGGAGCATCTGCAGGTCAAGGGCCGAGGCTTGA 309
QY 299 AGCTCATAGCCAGGAACAGGGTCAACCCAGACTGGGTGTGAGTGTGAAGATGTCCTTGA 358
Db 310 AGCTCATAGCCAGGAACAGGGTCAACCCAGACTGGGTGTGAGTGTGAAGATGTCCTTGA 369
QY 359 TGGGCGAGGATGACCCGCCAAATCCAGAGGAGGTGAAAACGCTGAAGAGCTGAAAA 418
Db 370 TGGGCGAGGATGACCCGCCAAATCCAGAGGAGGTGAAAACGCTGAAGAGCTGAAAA 429
QY 419 GCAATCAGCTGTTAAAGAGACACAGCTTGAATGATGACAGGCTGTCTTATGTTGAAA 478
Db 430 GCAATCAGCTGTTAAAGAGACACAGCTTGAATGATGACAGGCTGTCTTATGTTGAAA 489
QY 479 TTTGTTCA-TTAAATTTCTCCCAATAAAGCTTTACAGCCTTTCGCAAGAAA 528
Db 490 ATTCTTCATTGAAGTTCTCCCAATAAAGCTTTACAGCCTTTCGCAAGAAA 540

```

```

RESULT 9
LOCUS AA447559/c 489 bp mRNA linear EST 04-JUN-1997
DEFINITION zw81ell.sl Soares testis NHT Homo sapiens cDNA clone IMAGE:782636
3' similar to TR:G914905 G914905 GAGE-4 PROTEIN. [1] ;, mRNA
sequence.
ACCESSION AA447559
VERSION AA447559.1 GI:2161229
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1..(bases 1 to 489)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,
Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
,T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estowatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -4im13 fwd. Et from Amersham
High quality sequence stop: 427.
FEATURES
Location/Qualifiers
1..489
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:782636"

```


/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"

/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, inc., and primed with a Not I - oligo(dT) primer [5'.
TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization to Cor5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 101 a 139 c 105 g 144 t

Query Match 84.1%; Score 452.4; DB 9; Length 489;
Best Local Similarity 98.0%; Pred. No. 1.3e-100;
Matches 480; Conservative 0; Mismatches 6; Indels 4; Gaps 2;

QY 43 CCGGACTCTTTTCTCTACTGAGATTTCATCTGTGTGAATATGAGTTGGCGAGGAAGAT 102
|||||
Db 489 CCGGACTCTTTTCTCTACTGAGATTTCATCTGTGTGAATATGAGTTGGCGAGGAAGAT 430
|||||
QY 103 CGACC-----TATCGGCTAGACCAAGCGCTACGTAGAGCTCTCTGAATGATTGGGCTA 159
|||||
Db 429 CGACCTATTATTGGCCTAGACCAAGCGCTATGTACAGCTCTCTGAAGTGATT-GGCCTA 371
|||||
QY 160 TCGGGCCGAGCAGTTTCAGTGATGAAGTGAACCCAGCAACCTGAGAGGGGAACGAG 219
|||||
Db 370 TCGGGCCGAGCAGTTTCAGTGATGAAGTGAACCCAGCAACCTGAGAGGGGAACGAG 311
|||||
QY 220 CAACCTCAACGTCAGGATCCTGAGCTGCTCAGGAGGAGAGGATGAGGAGCATCTGCAG 279
|||||
Db 310 CAACCTCAACGTCAGGATCCTGAGCTGCTCAGGAGGAGAGGATGAGGAGCATCTGCAG 251
|||||
QY 280 GTCAAGGGCCGAAGCCTGAAGCTCTATAGCCAGAAACAGGGTCAACCAACAGACTGGGTGTG 339
|||||
Db 250 GTCAAGGGCCGAAGCCTGAAGCTCTATAGCCAGAAACAGGGTCAACCAACAGACTGGGTGTG 191
|||||
QY 340 AGTGTGAAGTGTCTCATGTTGGCGAGGAGATGGACCGCCCAATCCAGAGAGGTGAAGA 399
|||||
Db 190 AGTGTGAAGTGTCTCATGTTGGCGAGGAGATGGACCGCCCAATCCAGAGAGGTGAAGA 131
|||||
QY 400 CGCCTGAAGAGGTGAAGAAGCAATCACAGTGTAAAAAGAGACACAGTTTGAATATGATGCAG 459
|||||
Db 130 CGCCTGAAGAGGTGAAGAAGCAATCACAGTGTAAAAAGAGACACAGTTTGAATATGATGCAG 71
|||||
QY 460 GCTGCTCCTATGTTGGAAATTTGTTTCATTAAATTTCTCCCAATAAAGCTTTACAGCCTTC 519
|||||
Db 70 GCTGCTCCTATGTTGGAAATTTGTTTCATTAAATTTCTCCCAATAAAGCTTTACAGCCTTC 11
|||||
QY 520 TGCAGAGAA 529
|||||
Db 10 TGCAAGAGAA 1

RESULT 10
AW510753/c AW510753 457 bp mRNA linear EST 03-MAR-2000
LOCUS hd39405.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2911881 3' similar to SW:GGE4_HUMAN Q13068 GAGE-4 PROTEIN. ;,
mRNA sequence.

ACCESSION AW510753
VERSION AW510753.1 GI:7148831
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 457)
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps@mail.nih.gov
This clone is available royalty-free through LML ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Glibco.
Location/Qualifiers
1. .457
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:2911881"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP GCBI) were mixed, and as circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "

BASE COUNT 87 a 133 c 95 g 142 t
ORIGIN

Query Match 80.2%; Score 431.4; DB 10; Length 457;
Best Local Similarity 98.0%; Pred. No. 1.8e-95;
Matches 448; Conservative 0; Mismatches 6; Indels 3; Gaps 1;

QY 77 GTGAAATATGAGTTGGCGAGGAAGATCGACCTATC---GGCTAGACCAAGACGCTACGT 133
|||||
Db 457 GTGAAATATGAGTTGGCGAGGAAGATCGACCTATTTATTGGCCTAGACCAAGCGCGCTATGT 398
|||||
QY 134 AGAGCCTCTCTGAAATGATTGGGCTTATCGGCCCGAGCAGTTTCAGTGTGAAGTGGAAACC 193
|||||
Db 397 ACAGCCTCTCTGAAATGATTGGGCTTATCGGCCCGAGCAGTTTCAGTGTGAAGTGGAAACC 338
|||||
QY 194 AGCAACCTGTGAAGAGGGGAACAGCAACTCAACCTCAGGATCTTCAGCTGCAGCA 253
|||||
Db 337 AGCAACCTGTGAAGAGGGGAACAGCAACTCAACCTCAGGATCTTCAGCTGCAGCA 278
|||||
QY 254 GCGAGAGGATGAGGAGGAGCATCTGCAGGTCAAGGGCCGAGCCTGAAGCTCATAGCCAGGA 313
|||||
Db 277 GCGAGAGGATGAGGAGGAGCATCTGCAGGTCAAGGGCCGAGCCTGAAGCTCATAGCCAGGA 218
|||||
QY 314 ACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTCTGATGGCGAGGAGATGGA 373
|||||
Db 217 ACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTCTGATGGCGAGGAGATGGA 158
|||||
QY 374 CCGCCCAATCCAGAGGAGGTGAAAACCGCTGAAGAGGTGAAAAGCAATCACAGTGTGA 433
|||||
Db 157 CCGCCCAATCCAGAGGAGGTGAAAACCGCTGAAGAGGTGAAAAGCAATCACAGTGTGA 98
|||||
QY 434 AAAGAAGACACGTTGAAATGATGCAGGCTGCTCTATGTTGGAAATTTGTTTCATTAAAT 493
|||||
Db 97 AAAGAAGACACGTTGAAATGATGCAGGCTGCTCTATGTTGGAAATTTGTTTCATTAAAT 38
|||||
QY 494 TCTCCCAATAAAGCTTTTACAGCTTCTCTGCAAAAGAAAA 530
|||||
Db 37 TCTCCCAATAAAGCTTTTACAGCTTCTGCAAAAGAAAA 1
|||||

RESULT 11
AI381509/c AI381509 447 bp mRNA linear EST 28-MAR-1999
LOCUS te76B07.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2092597 3' similar to TR:Q13070 Q13070 GAGE-6 PROTEIN. ;,
mRNA sequence.
ACCESSION AI381509

```
VERSION AI381509.1 GI:4194290
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 497 Std Error: 0.00
Seq primer: -40UP from Gibco.
FEATURES
source
1. .447
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2092597"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CCGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729036-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "
BASE COUNT 84 a 130 c 95 g 138 t
ORIGIN
Query Match 78.9%; Score 424.6; DB 9; Length 447;
Best Local Similarity 98.4%; Pred. No. 8.5e-94;
Matches 440; Conservative 0; Mismatches 4; Indels 3; Gaps 1;
QY 88 GTTGGCGAGGAAGATCGACCC---TATCGCGCTAGACCAAGAGCGTACGTAGAGCCTCCTG 144
DB 447 GTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGCGCTATGTACAGCCTCCTG 388
QY 145 AATGATTGGCGCTATCGGCGCCGAGCAGTTTCAGTGTGAAGTGGAAACGAGCAACACCTG 204
DB 387 AATGATTGGCGCTATCGGCGCCGAGCAGTTTCAGTGTGAAGTGGAAACGAGCAACACCTG 328
QY 205 AAGAAGGGGAACAGCAACTCAACGTCTAGGATCCTCGACGTCTCAGGAGGGAGGATG 264
DB 327 AAGAAGGGGAACAGCAACTCAACGTCTAGGATCCTCGACGTCTCAGGAGGGAGGATG 268
QY 265 AGGAGGATCTGAGGTCAGAGGCGGAGACCTGGAAGCTCATAGCCAGGAACAGGGTCACC 324
DB 267 AGGAGGATCTGAGGTCAGAGGCGGAGACCTGGAAGCTCATAGCCAGGAACAGGGTCACC 208
QY 325 CACAGACTGGGTGTGAAGTGTCTGTGATGGGAGGAGATGGAACCCGCCCAATC 384
DB 207 CACAGACTGGGTGTGAAGTGTCTGTGATGGGAGGAGATGGAACCCGCCCAATC 148
QY 385 CAGAGAGGTGAAGACCCCTGAAGAGGTGAAGCAATCACAGTGTGTAAGAGAGACAC 444
DB 147 CAGAGAGGTGAAGACCCCTGAAGAGGTGAAGCAATCACAGTGTGTAAGAGAGACAC 88
QY 445 GTTGAATGATGAGGCTGTCTTATGTTGGAATTTGTTTCAATTAATTTCTCCCAATA 504
DB 87 GTTGAATGATGAGGCTGTCTTATGTTGGAATTTGTTTCAATTAATTTCTCCCAATA 28
QY 508 AGCTTTACAGCCTTCTGCAAGAAAAA 531
|||||
```

```
DB 27 AGCTTTACAGCCTTCTGCAAGAAAAA 1
RESULT 12
AA738037 464 bp mRNA linear EST 22-JAN-1998
LOCUS nx15e11.s1 NCI CGAP GC3 Homo sapiens cDNA clone IMAGE:1256204 3,
DEFINITION similar to TR:Q13070 Q13070 GAGE-6 PROTEIN. ;, mRNA sequence.
ACCESSION AA738037
VERSION AA738037.1 GI:2768794
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 511 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 435.
FEATURES
source
1. .464
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1256204"
/clone_lib="NCI CGAP GC3"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 86 a 132 c 93 g 153 t
ORIGIN
Query Match 78.9%; Score 424.4; DB 9; Length 464;
Best Local Similarity 96.9%; Pred. No. 9.5e-94;
Matches 444; Conservative 0; Mismatches 11; Indels 3; Gaps 1;
QY 84 ATGAGTTGCGAGGAAGATCGACC---TATCGGCTAGACCAAGAGCGTACGTAGAGCCT 140
DB 464 ATGAGTTGCGAGGAAGATCGACCTGTTATTGGCTAGTCCAAGAGCGTATGTACAGCCT 405
QY 141 CCTGAAATGATTGGGCTTATCGGCGCCGAGCAGTTTCAGTGTGAAGTGAAGCAACAGCAACA 200
DB 404 CCTGAAATGATTGGGCTTATCGAGCCGAGCAGTTTCAGTGTGAAGTGAAGCAACAGCAACA 345
QY 201 CCTGAAGAGGGGAACAGCAACTCAACGTTCAGGATCTTCGAGCTGCTCAGGAGGGAGAG 260
DB 344 CCTGAAGAGGGGAACAGCAACTCAACGTTCAGGATCTTCGAGCTGCTCAGGAGGGAGAG 285
QY 261 GATGAGGGAGCATCTCAGGTCAGGCGCCGAGCGCTGAAGCTCATAGCCAGGAACAGGGGT 320
DB 284 GATGAGGGAGCATCTCAGGTCAGGCGCCGAGCGCTGAAGCTCATAGCCAGGAACAGGGGT 225
QY 321 CACCACAGACTGGGTGTGAGTGTGAAGATGTCCTGATGGGAGGAGATGACCCGCCA 380
|||||
```

Db 224 CACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTGTATGGCAGGAGATGGACCCGCA 165
Qy 381 AATCCAGAGAGGTGAAAAACGCTGAAGAGAGGTGAAAGCAATCACAGTGTAAAAAGAG 440
Db 164 AATCCAGAGAGGTGAAAAACGCTGAAGAGAGGTGAAAGCAATCACAGTGTAAAAAGAG 105
Qy 441 ACAGCTTGAATGATGACAGCTGCTCTATGTTGGAATTTGTTCAATTAATTTCTCCCA 500
Db 104 GCAGCTTGAATGATGACAGCTGCTCTATGTTGGAATTTGTTCAATTAATTTCTCCCA 45
Qy 501 ATAAAGCTTTACAGCTTCTGCAAGAAAAA 538
Db 44 ATAAGAGTTTACAGCTTCTGCAAAAAA 7
RESULT 13
BG700088 581 bp mRNA linear EST 07-MAY-2001
LOCUS 602679431F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4812462 5',
DEFINITION mRNA sequence.
ACCESSION BG700088
VERSION BG700088.1 GI:13969078
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 581)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAMI0704 row: m column: 07
High quality sequence stop: 548.
FEATURES
source
1..581
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4812462"
/tissue_type="NIH_MGC_95"
/lab_host="hippoCampus"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTWN-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 164 a 142 c 166 g 109 t
ORIGIN
Query Match 78.8%; Score 424; DB 12; Length 581;
Best local Similarity 92.8%; Pred. No. 1.2e-93;
Matches 501; Conservative 0; Mismatches 30; Indels 9; Gaps 5;
Qy 8 GGAGCTGTGAGGACGTGCTGTGGTCTCGCGTCCGGACTCTTTTCTCTACTGAGA 67
Db 25 GGAGCTGTGAGGACGTGCTGTGGTCTCGCGTCCGGACTCTTTTCTCTACTGAGA 84
Qy 68 TTCATCTGTGAAATATAGTTGGCGGAGAAAGATCGACCTATCGGCCTAGACCAAGACG 127

Db 85 TTCACTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGACCAAGACG 144
Qy 128 CTAGGTAGAGCCTCTCTGA-BATGATTGGGCCCTATCGGCCCGGAGAGTTTCAGTGATGAAG 186
Db 145 CTAGGTAGAGCCTCTCTGACAAATGATTGGGCCCTATCGGCCCGGAGCAAGTTTCAGTGATGAAG 204
Qy 187 TGGAAACAGCAACACCTTGAAGAGGGGAACCCAGCAACTCAACGTCAGGATCCTTGCAGCTG 246
Db 205 TGGAAACAGCAACACCTTGAAGAGGGGAACCCAGCAACTCAACGTCAGGATCCTTGCAGCTG 264
Qy 247 CTCAGGAGGAG-AGGATGAGGGAGCATCTGCAGTCAAGGCCCGAGGCTGAAGCTCAT 305
Db 265 CTCAGGAGGAGGAGGCTGAGGGAGCATCTGCAGGTCAGGGCCGAGGCTGAAGCTGAT 324
Qy 306 AGCCAGGAACAGGGTCAACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTCTGATGGGCAG 365
Db 325 AGCCAGGAACAGGGTCAACCCACAGACTGGGTGTGAGTGTGAAGATGGTCTCTGATGGGCAG 384
Qy 366 GAGATGACCCGCCCAATCCAGAGGAGGTGAAAAACGCTTGAAGA-AGGTGAAAAGCAATC 424
Db 385 GAGATGACCCGCCCAATCCAGAGGAGCGTGAAGAACGCTTGAAGACAGGTGAAAAAGCAATC 444
Qy 425 ACA---GTGTTAAAGAAAGACACGTTGAAATGATGCAGGCTGCTCTATCTTGGAAATTT 481
Db 445 ACAAGTGTGTACAAAGAGAGGACCGTTGACATGATGCAGGCTGCTCTATCTTGGAAACTC 504
Qy 482 GTTCA---TTAAATTTCTCCCAATAAAGCTTTTACAGCCTTCTGCAAGAAAAA 538
Db 505 TGTAACTTACATATCTCCCACTAAAGCTATACAGCCTTCTGCAACAAACACCAACA 564
RESULT 14
AW016546/c 418 bp mRNA linear EST 10-SEP-1999
LOCUS AW016546
DEFINITION UI-H-BI0p-abg-g-06-0-UI.s1 NCI_CGAP_Sub2 Homo sapiens cDNA clone
IMAGE:2711986 3', mRNA sequence.
ACCESSION AW016546
VERSION AW016546
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 418)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes
FEATURES
source
Location/Qualifiers
1..418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2711986"
/lab_host="NCI_CGAP_Sub2"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The
NCI_CGAP Sub2 library is a subtracted library derived from
BI. BI constitutes a mixture of 21 normalized or
subtracted NCI_CGAP libraries: NCI_CGAP_Co4, NCI_CGAP_Pr22
, NCI_CGAP_Pr28, NCI_CGAP_Co10, NCI_CGAP_Co16,

NCI CGAP Kid5, NCI CGAP Kid12, NCI CGAP Kid3,
NCI CGAP Kid11, NCI CGAP Lym2, NCI CGAP Br2, NCI CGAP Co8,
NCI CGAP CLL1, NCI CGAP Ie12, NCI CGAP Brn23, NCI CGAP Lu5
, NCI CGAP Lu24, NCI CGAP Lu19, NCI CGAP GC4, NCI CGAP GC6
, NCI CGAP Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with a
driver whose composition is detailed below: NCI CGAP Kid3
pool 1 LLAM 3334-3337, 3682-3683, 3798-3803 (IMAGE_Kid3
CloneIDs 1322376-1323911, 1456008-1456775, 1500552-1502855
) NCI CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725,
3776-3778 (IMAGE CloneIDs 1323912-1325831,
1471368-1472903, 1492104-1493255) NCI CGAP Lu5 pool 1 LLAM
3575-3582, 3851-3854 (IMAGE CloneIDs 1414920-1417991,
1520904-1522439) NCI CGAP GC4 pool 1 LLAM 3164-3167,
3716-3720, 3733-3735 (IMAGE CloneIDs 1257096-1258631,
1469064-1470983, 1475592-1476743) NCI CGAP Pr22 pool 1
LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE CloneIDs
985608-986759, 1101192-1101959, 1217928-1220615)
NCI CGAP Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE
CloneIDs 1057416-1061255, 1144584-1145351) The resulting
subtracted library contained 4 million recombinants.
Subtraction was performed as previously described [Bonaldo
, Lennon & Soares (1996): Normalization and Subtraction:
Two Approaches To Facilitate Gene Discovery. Genome
Research 6, 791-806.
TAG LIB=NCI CGAP GC4
TAG_TISSUE=germ Cell
TAG_SEQ=AAATC"

BASE COUNT 74 a 119 c 87 g 138 t

Query Match 76.9%; Score 413.8; DB 10; Length 418;
Best Local Similarity 99.5%; Pred. No. 3.8e-91;
Matches 415; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 122 AAGACGCTAGTAGAGCTCTGAAATGATTGGGCTTATGGCCCGAGAGTTCAGTGA 181
DB 418 AAGACGCTAGTAGAGCTCTGAAATGATTGGGCTTATGGCCCGAGAGTTCAGTGA 359
QY 182 TGAAGTGGAAACAGCAACACCTGAAGAGGGGAACAGCAACTCAAGTCAAGTCTGTC 241
DB 358 TGAAGTGGAAACAGCAACACCTGAAGAGGGGAACAGCAACTCAAGTCAAGTCTGTC 299
QY 242 AGCTGCTCAGAGGAGGAGGATGAGGAGCATCTGCAGGTCAAGGCGGAGGCTGAAGC 301
DB 298 AGCTGCTCAGAGGAGGAGGATGAGGAGCATCTGCAGGTCAAGGCGGAGGCTGAAGC 239
QY 302 TCATAGCCGGAACAGGCTACCCACAGACTGGGTGTGAGTGAAGATGGTCTGATGG 361
DB 238 TCATAGCCGGAACAGGCTACCCACAGACTGGGTGTGAGTGAAGATGGTCTGATGG 179
QY 362 GCAGGAGATGGACCCGCAAAATCCAGAGAGGTGAAACGCCCTGAAGAGGTGAAAGCA 421
DB 178 GCAGGAGATGGACCCGCAAAATCCAGAGAGGTGAAACGCCCTGAAGAGGTGAAAGCA 119
QY 422 ATCAGAGTGTAAAGAGACACGTTGAAATGATCAGGCTGCTCTATGTTGAAATTT 481
DB 118 ATCAGAGTGTAAAGAGACACGTTGAAATGATCAGGCTGCTCTATGTTGAAATTT 59
QY 482 GTTCATTAAATTTCTCCCAATAAAGCTTTACAGCTTCTGCAAGAAAAA 538
DB 58 GTTCATTAAATTTCTCCCAATAAAGCTTTACAGCTTCTGCAAGAAAAA 2

RESULT 15
BG120336
LOCUS BG120336 480 bp mRNA linear EST 30-JAN-2001
DEFINITION 602353732F1 NIH_MGC_90 Homo sapiens cdna clone IMAGE:4451958 5',
mRNA sequence.
ACCESSION BG120336
VERSION BG120336.1 GI:12613845
KEYWORDS EST.

SOURCE
ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 480)

AUTHORS

NIH-MGC http://mgc.nci.nih.gov/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cdna Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM10239 row: p column: 07

High quality sequence stop: 480.

FEATURES

source

Location/Qualifiers

1..480

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4451958"

/clone_lib="NIH_MGC_90"

/tissue_type="adenocarcinoma, cell_line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: liver; Vector: pCMV-Sport6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 1.7 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC Library."

BASE COUNT 170 a 93 c 133 g 84 t

ORIGIN

Query Match 76.8%; Score 413.4; DB 12; Length 480;

Best Local Similarity 97.4%; Pred. No. 4.8e-91;

Matches 420; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 108 TATCGCGCTAGACCAAGACGCTAGTAGAGCTCTCTGAAATGATTGGGCTTATGGCGCC 167

DB 15 TATTGGCGCTAGACCAAGCGCTATGTACAGCTCTCTGAAATGATTGGGCTTATGGCGCC 74

QY 168 GAGCAGTTCAGTGATGAAGTGGAAACACCTGAAAGAGGGGAACCCAGCAACTCAA 227

DB 75 GAGCAGTTCAGTGATGAAGTGGAAACACCTGAAAGAGGGGAACCCAGCAACTCAA 134

QY 228 CGTCAGGATCCTGCAGCTGCTCAGGAGGAGGATGAGGGAGCATCTGCAGGTCAAGGG 287

DB 135 CGTCAGGATCCTGCAGCTGCTCAGGAGGAGGATGAGGGAGCATCTGCAGGTCAAGGG 194

QY 288 CCGAAGCCTGAAGCTCATAGCCAGGACAGGGTCAACCCAGACTGGGTGTGAGTGTGAA 347

DB 195 CCGAAGCCTGAAGCTCATAGCCAGGACAGGGTCAACCCAGACTGGGTGTGAGTGTGAA 254

QY 348 GATGTCCTGATGGGAGGAGATGGACCCGCAAAATCCAGAGAGGTGAAACGCCCTGAA 407

DB 255 GATGTCCTGATGGGAGGAGATGGACCCGCAAAATCCAGAGAGGTGAAACGCCCTGAA 314

QY 408 GAAGGTGAAAGCAATCAGTGTAAAAAGAAACACGTTGAAATGATCAGGCTGCTCC 467

DB 315 GAAGGTGAAAGCAATCAGTGTAAAAAGAAAGGACCGTTGAAATGATCAGGCTGCTCC 374

QY 468 TATGTTGAAATTTGTTTCAATTAATTTCTCCCAATAAAGCTTTACAGCTTCTGCAAGA 527

DB 375 TATGTTGAAATTTGTTTCAATTAATTTCTCCCAATAAAGCTTTACAGCTTCTGCAAGA 434

QY 528 AAAAAAAAAA 538

DB 435 AAAAAAAAAA 445

Search completed: April 9, 2003, 17:06:32

Wed Apr 9 18:05:23 2003

us-09-782-745-14.rst

Page 11

Job time : 1535 secs



GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2003, 10:55:39 ; Search time 224 Seconds
(without alignments)
5408.817 Million cell updates/sec

Title: US-09-782-745-14
Perfect score: 538
Sequence: 1 ACGCCAGGAGCTGTGAGGC.....CTGCAAGAAAAA 538

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SID82/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SID82/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SID82/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SID82/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SID82/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SID82/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SID82/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SID82/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SID82/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SID82/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SID82/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SID82/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SID82/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SID82/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SID82/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SID82/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SID82/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SID82/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SID82/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SID82/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SID82/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SID82/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SID82/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SID82/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	538	100.0	538	20 AAX90519	GAGE-2 tumour reje
2	531.6	98.8	1245	22 AAS60826	Human cancer agent
3	512.8	95.3	540	20 AAX90521	GAGE-4 tumour reje
4	510.2	94.8	540	20 AAX90523	GAGE-6 tumour reje
5	504.8	93.8	532	20 AAX90522	GAGE-5 tumour reje
6	501.4	93.2	528	24 ABL66323	Lung cancer relate
7	500.6	93.0	539	19 AAV18721	CDNA encoding GAGE
8	496.8	92.3	532	19 AAV18720	CDNA encoding GAGE
9	493	91.6	535	19 AAV18717	CDNA encoding GAGE

10	490.6	91.2	541	19 AAV18719	CDNA encoding GAGE
11	429.2	79.8	560	20 AAX90520	GAGE-3 tumour reje
12	421.2	78.3	560	19 AAV18718	CDNA encoding GAGE
13	399.4	74.2	1024	21 AAZ97217	Human prostate can
14	397.4	73.9	1024	21 AAZ97216	Human prostate can
15	374.4	69.6	646	17 AAT31861	GAGE tumour reject
16	374.4	69.6	646	20 AAX90518	GAGE tumour reject
17	368	68.4	646	19 AAV05540	CDNA encoding GAGE
18	366.4	68.1	648	16 AAO81405	GAGE tumour reject
19	357.8	66.5	430	21 AAC02129	Human secreted pro
20	301.4	56.0	530	22 AAS60104	Human cancer agent
21	299.2	55.6	365	22 AAS60496	Human cancer agent
22	203.6	37.8	650	21 AAS9322	CDNA encoding a ca
23	194	36.1	532	22 AAI60530	Human polynucleoti
24	189.2	35.2	659	22 AAI58744	Human polynucleoti
25	183.8	34.2	642	24 ABA93876	Human G protein-co
26	181.8	33.8	673	22 ABA93198	Human polynucleoti
27	173.8	32.3	611	22 AAD14983	Human NOV4 DNA. H
28	172.4	32.0	503	23 AAS69484	DNA encoding novel
29	165.6	30.8	656	21 AAZ97402	Human prostate can
30	156	29.0	580	22 AAF59637	Human cell cycle a
31	139.8	26.0	661	23 AAS69486	DNA encoding novel
32	131.4	24.4	11462	19 AAV05532	Nucleic acid seque
33	130.4	24.2	475	22 AAD14981	Human NOV2 DNA. H
34	127.8	23.8	665	22 AAK52902	Human polynucleoti
35	120.8	22.5	1051	22 AAD14982	Human NOV3 DNA. H
36	111.4	20.7	637	24 ABK50912	XAGE-1 gene encodi
37	111.4	20.7	762	22 AAH64751	Human secreted pro
38	108.4	20.1	781	22 AAF68861	Human lung tumour
39	108.4	20.1	781	24 ABK38772	CDNA encoding L552
40	108	20.1	399	22 AAF68151	Human lung tumour
41	108	20.1	399	24 ABK38062	CDNA encoding clon
42	105.4	19.6	750	23 AAS91235	DNA encoding novel
43	105.2	19.6	571	21 AAZ33316	Human secreted pro
44	104.4	19.4	121724	24 ABQ88143	Human osteoblast d
45	103.6	19.3	505	24 AAD24228	Human differential

ALIGNMENTS

RESULT 1
AAX90519
ID AAX90519 standard; cDNA; 538 BP.
XX
AC AAX90519;
XX
DT 30-SEP-1999 (first entry)
XX
GAGE-2 tumour rejection antigen clone nucleotide sequence.
DE
DE DE DE
KW Human leukocyte antigen; HLA-A29; tumour rejection antigen;
KW detection; therapy; pathological condition; cancer; CTL;
KW cytolytic T lymphocyte; GAGE; ss.
XX
OS Homo sapiens.
OS
PN XX
PN WO9937665-A1.
XX
PD 29-JUL-1999.
XX
PF 12-JAN-1999; 99WO-US00775.
XX
PR 23-JAN-1998; 98US-0012818.
XX
PA (LUDM-) LUDWIG INST CANCER RES.
XX
PI Boon-Falleur T, Debacker O, Van Den Eynde B, Van Der Bruggen P;
XX
XX WPI; 1999-469111/39.
XX
PT New isolated peptides which bind to HLA-A29 molecules, which are
PT tumour rejection antigens used for detection and therapy of

PT pathological conditions, e.g. cancer
PS Example 13; Fig 4; 62pp; English.
XX
CC The present invention describes peptides which bind to human leukocyte
antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into
tumour rejection antigens. They can be used for detecting cytolytic T
lymphocytes (CTLs) in pathological conditions such as cancer and in
HLA-typing assays. Complexes of HLA-29 molecules and the peptides can
be used for stimulating CTLs in vivo. The present sequence represents
a GAGE tumour rejection antigen clone, from an example from the present
invention.
XX
SQ Sequence 538 BP; 160 A; 116 C; 155 G; 107 T; 0 other;
Query Match 100.0%; Score 538; DB 20; Length 538;
Best Local Similarity 100.0%; Pred. No. 3.9e-146; Mismatches 0; Indels 0; Gaps 0;
Matches 538; Conservative 0;
QY 1 ACGCCAGGGAGCTGTGAGGCACTGCTGTGTGTTCTCTGCGTCCGCACTCTTTTCTCTCT 60
DB 1 ACGCCAGGGAGCTGTGAGGCACTGCTGTGTGTTCTCTGCGTCCGCACTCTTTTCTCTCT 60
QY 61 ACTGAGATTCATCTGTGTGAATATAGTTGGCGAGGAAGATCGACTATCGGCCTAGAC 120
DB 61 ACTGAGATTCATCTGTGTGAATATAGTTGGCGAGGAAGATCGACTATCGGCCTAGAC 120
QY 121 CAAGAGCTACGTAGAGCTCTGAAATGATTGGGCTATGCGGCCGAGCAGTTCAGTG 180
DB 121 CAAGAGCTACGTAGAGCTCTGAAATGATTGGGCTATGCGGCCGAGCAGTTCAGTG 180
QY 181 ATCAAGTGGAAACCAACCACTGAAGAGGGGGAACCACTCAACGTGAGGATCCTG 240
DB 181 ATCAAGTGGAAACCAACCACTGAAGAGGGGGAACCACTCAACGTGAGGATCCTG 240
QY 241 CAGCTGCTCAGGAGGAGAGATGAGGAGCATCTGCAAGTCAAGGGCCGAAGCCTGAAG 300
DB 241 CAGCTGCTCAGGAGGAGAGATGAGGAGCATCTGCAAGTCAAGGGCCGAAGCCTGAAG 300
QY 301 CTCATGCCAGGAACAGGGTCAACCAACAGCTGGGTGTGAGTGTGAAGTGTCTCTGATG 360
DB 301 CTCATGCCAGGAACAGGGTCAACCAACAGCTGGGTGTGAGTGTGAAGTGTCTCTGATG 360
QY 361 GGCAGGAGATGGACCGCCAAATCCAGAGGAGGTGAAAACCGCTGAAAGAGGTGAAAAGC 420
DB 361 GGCAGGAGATGGACCGCCAAATCCAGAGGAGGTGAAAACCGCTGAAAGAGGTGAAAAGC 420
QY 421 AATCAGATGTTAAAGAAAGACACCTTGAATGATGACGCTGCTCTATGTTGGAATTT 480
DB 421 AATCAGATGTTAAAGAAAGACACCTTGAATGATGACGCTGCTCTATGTTGGAATTT 480
QY 481 TGTTCATTAATAATCTCCCAATAAGCTTTACAGCCTTCTGCAAGAAAGAAAAA 538
DB 481 TGTTCATTAATAATCTCCCAATAAGCTTTACAGCCTTCTGCAAGAAAGAAAAA 538
RESULT 2
AAS60826
ID AAS60826 standard; cDNA; 1245 BP.
XX
AC AAS60826;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human cancer agent-sensitive marker #328.
XX
KW Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
KW squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
KW lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
KW Hodgkin's disease; glioma; ss.
XX
OS Homo sapiens.
XX

PN WO200179556-A2.
XX
PD 25-OCT-2001.
XX
PF 13-APR-2001; 2001WO-US12132.
XX
PR 14-APR-2000; 2000US-197538P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Brown JL, Bolt A, Van Huffel C;
XX
DR WPI; 2001-602933/68.
XX
PT Novel nucleic acid, used as a marker to determine the effectiveness of
using TAXOL to treat cancer cell growth in individuals -
XX
PS Claim 1; Page 340; 527pp; English.
XX
CC The invention relates to 1046 novel nucleic acids which are used as
markers for determining the sensitivity of a cancer cell to the
anticancer agent TAXOL. Cancer cells can be treated with TAXOL when
they are shown to express one of the 242 sensitivity markers or the
cells are shown not to express one of the 804 resistance markers.
CC The methods can be used to determine the effectiveness of TAXOL.
CC in the treatment of cancer cell growth in an individual. The markers
can be used as targets in developing anti-cancer agents such as
chemotherapeutic compounds. The markers can also be used as targets in
displaying resistance to agents and exhibit expression of the markers. The
anticancer agents developed by the novel method can be used to treat
cancer. Probes based on the markers can be used to detect transcripts or
genomic sequences corresponding to the markers, in the identification of
cells or tissues which mis-express the protein. Cancers which may
be targeted include carcinoma (e.g. squamous cell carcinoma),
sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic leukaemia),
lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's disease and
tumours (e.g. glioma). The present sequence is one of the 1046
novel cancer cell markers.
XX
SQ Sequence 1245 BP; 382 A; 308 C; 390 G; 153 T; 12 other;
Query Match 98.8%; Score 531.6; DB 22; Length 1245;
Best Local Similarity 99.3%; Pred. No. 4.2e-144;
Matches 534; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 AGCCAGGGAGCTGTGAGGCACTGCTGTGTGTTCTCTGCGTCCGCACTCTTTTCTCTCT 60
DB 179 AGCCAGGGAGCTGTGAGGCACTGCTGTGTGTTCTCTGCGTCCGCACTCTTTTCTCTCT 238
QY 61 ACTGAGATTCATCTGTGTGAATATAGTTGGCGAGGAAGATCGACTATTCGCTAGAC 120
DB 239 ACTGAGATTCATCTGTGTGAATATAGTTGGCGAGGAAGATCGACTATTCGCTAGAC 298
QY 121 CAAGAGCCTACGTAGAGCCTCTGAAATGATTGGGCCCTATGCGGCCGAGCAGTTCAGTG 180
DB 299 CAAGAGCCTACGTAGAGCCTCTGAAATGATTGGGCCCTATGCGGCCGAGCAGTTCAGTG 358
QY 181 ATGAAGTGGAAACCAACCACTGAAAGAGGGGAACCACTCAACGTGAGGATCCTG 240
DB 359 ATGAAGTGGAAACCAACCACTGAAAGAGGGGAACCACTCAACGTGAGGATCCTG 418
QY 241 CAGCTGCTCAGGAGGAGAGGATGAGGAGCATCTGCAAGTCAAGGGCCGAAGCCTGAAG 300
DB 419 CAGCTGCTCAGGAGGAGAGGATGAGGAGCATCTGCAAGTCAAGGGCCGAAGCCTGAAG 478
QY 301 CTCATGCCAGGAACAGGGTCAACCAACAGCTGGGTGTGAGTGTGAAGATGTCCTGATG 360
DB 479 CTCATGCCAGGAACAGGGTCAACCAACAGCTGGGTGTGAGTGTGAAGATGTCCTGATG 538
QY 361 GGCAGGAGATGGACCGCCCAATCCAGAGGAGGTGAAACCGCTGAAAGAGGTGAAAAGC 420
DB 539 GGCAGGAGATGGACCGCCCAATCCAGAGGAGGTGAAACCGCTGAAAGAGGTGAAAAGC 598

QY 421 AATCACAGTGTAAAGAAAGACACGTTGAAATGATGAGGCTGCTCTATGTTGGAATT 480
DB 599 AATCACAGTGTAAAGAAAGACGTTGAAATGATGAGGCTGCTCTATGTTGGAATT 658
QY 481 TGTTCATTAAATTTCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 538
DB 659 TGTTCATTAAATTTCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 716

RESULT 3
AAAX90521
ID AAAX90521 standard; cDNA; 540 BP.
XX
AC AAAX90521;
XX
DT 30-SEP-1999 (first entry)
XX
DE GAGE-4 tumour rejection antigen clone nucleotide sequence.
XX
KW Human leukocyte antigen; HLA-A29; tumour rejection antigen;
KW detection; therapy; pathological condition; cancer; CTL;
KW cytolytic T lymphocyte; GAGE; ss.
XX
OS Homo sapiens.
XX
PN WO9937665-A1.
XX
PD 29-JUL-1999.
XX
PF 12-JAN-1999; 99WO-US00775.
XX
PR 23-JAN-1998; 98US-0012818.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Boon-Falleur T, Debacker O, Van Den Eynde B, Van Der Bruggen P;
XX WPI; 1999-469111/39.
XX
PT New isolated peptides which bind to HLA-A29 molecules, which are
PT tumour rejection antigens used for detection and therapy of
PT pathological conditions, e.g. cancer
XX
PS Example 13; Fig 4; 62pp; English.
XX
CC The present invention describes peptides which bind to human leukocyte
CC antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into
CC tumour rejection antigens. They can be used for detecting cytolytic T
CC lymphocytes (CTLs) in pathological conditions such as cancer and in
CC HLA-typing assays. Complexes of HLA-29 molecules and the peptides can
CC be used for stimulating CTLs in vivo. The present sequence represents
CC a GAGE tumour rejection antigen clone, from an example from the present
CC invention.
XX
SQ Sequence 540 BP; 159 A; 114 C; 156 G; 111 T; 0 other;
Query Match 95.3%; Score 512.8; DB 20; Length 540;
Best Local Similarity 98.1%; Pred. No. 8.2e-139;
Matches 530; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

QY 2 CGCCAGGAGCTGTGAGGAGCTGCTGTGTGTTCTCGCTCGGACTCTTTTCTCTA 61
DB 1 CGCCAGGAGCTGTGAGGAGCTGCTGTGTGTTCTCGCTCGGACTCTTTTCTCTA 60
QY 62 CTGAGATTCACTGTGTGAATATGATGTCGCGAGGAATCGACC---TATCGGCCTAG 118
DB 61 CTGAGATTCACTGTGTGAATATGATGTCGCGAGGAATCGACCATTATTGCGCTAG 120
QY 119 ACCAAGAGCTACGTAGAGCCTCTCAATGATTGGCCCTATGCGCCGAGCAGTTCCAG 178
DB 121 ACCAAGGCGCTATGTACAGCCCTCTGAAATGATTGGGCCCTATGCGCCGAGCAGTTCCAG 180

QY 179 TGATGAAGTGAACACAGCAACACCTGTAAGAAAGGGGAAACAGCAACTCAACGTCAGGATCC 238
DB 181 TGATGAAGTGAACACAGCAACACCTGTAAGAAAGGGGAAACAGCAACTCAACGTCAGGATCC 240
QY 239 TGCAGCTGCTCAGGAGGAGGAGATGAGGAGCATCTGCAGGTCAAGGCCCGAAGCCTGA 298
DB 241 TGCAGCTGCTCAGGAGGAGGAGATGAGGAGCATCTGCAGGTCAAGGCCCGAAGCCTGA 300
QY 299 AGCTCATAGCCAGGAACAGGGGTCAACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 358
DB 301 AGCTCATAGCCAGGAACAGGGGTCAACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 360
QY 359 TGGCAGGAGATGGACCCGCAAAATCCAGAGAGGTGAACCGCTGAAGAGGTGA 418
DB 361 TGGCAGGAGATGGACCCGCAAAATCCAGAGAGGTGAACCGCTGAAGAGGTGA 420
QY 419 GCAATCAGAGTGTAAAGAAAGACACAGTTGAAATGATGCAGGCTGCTCTATGTTGAAA 478
DB 421 GCAATCAGAGTGTAAAGAAAGACAGTTGAAATGATGCAGGCTGCTCTATGTTGAAA 480
QY 479 TTTGTTTCAATAAAATTTCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 538
DB 481 TTTGTTTCAATAAAATTTCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 540

RESULT 4
AAAX90523
ID AAAX90523 standard; cDNA; 540 BP.
XX
AC AAAX90523;
XX
DT 30-SEP-1999 (first entry)
XX
DE GAGE-6 tumour rejection antigen clone nucleotide sequence.
XX
KW Human leukocyte antigen; HLA-A29; tumour rejection antigen;
KW detection; therapy; pathological condition; cancer; CTL;
KW cytolytic T lymphocyte; GAGE; ss.
XX
OS Homo sapiens.
XX
PN WO9937665-A1.
XX
PD 29-JUL-1999.
XX
PF 12-JAN-1999; 99WO-US00775.
XX
PR 23-JAN-1998; 98US-0012818.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Boon-Falleur T, Debacker O, Van Den Eynde B, Van Der Bruggen P;
XX WPI; 1999-469111/39.
XX
PT New isolated peptides which bind to HLA-A29 molecules, which are
PT tumour rejection antigens used for detection and therapy of
PT pathological conditions, e.g. cancer
XX
PS Example 13; Fig 4; 62pp; English.
XX
CC The present invention describes peptides which bind to human leukocyte
CC antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into
CC tumour rejection antigens. They can be used for detecting cytolytic T
CC lymphocytes (CTLs) in pathological conditions such as cancer and in
CC HLA-typing assays. Complexes of HLA-29 molecules and the peptides can
CC be used for stimulating CTLs in vivo. The present sequence represents
CC a GAGE tumour rejection antigen clone, from an example from the present
CC invention.
XX
SQ Sequence 540 BP; 159 A; 113 C; 157 G; 111 T; 0 other;
Query Match 94.8%; Score 510.2; DB 20; Length 540;

Best Local Similarity		98.0%;	Pred. No. 4.7e-138;
Matches 528; Conservative		0;	Mismatches 8; Indels 3; Gaps 1;
QY	3	GCACGGAGCTGTGAGGCAGTGCCTGTGTGGTTCCTGCCGCTCCGGACTCTTTTTCCTCTAC	62
Db	1	GCACGGAGCTGTGAGGCAGTGCCTGTGTGGTTCCTGCCGCTCCGGACTCTTTTTCCTCTAC	60
QY	63	TGAGATTTCATCTGTGTAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGA	119
Db	61	TGAGATTTCATCTGTGTAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGA	120
QY	120	CCAGACGCTACGTAGAGCCTCTGAAATGATTTGGCCCTATGCGGCCCGAGCAGTTTCAGT	179
Db	121	CCAGGCGCTATGTACAGCCTCTCTGAAGTGAATGGGCCCTATGCGGCCCGAGCAGTTTCAGT	180
QY	180	GATGAAGTGGAAACCCAGCAACACCTTGAAAGGGGAACACGAACTCAACGTCAGGATCCT	239
Db	181	GATGAAGTGGAAACCCAGCAACACCTTGAAAGGGGAACACGAACTCAACGTCAGGATCCT	240
QY	240	GCAGCTGCTCAGGAGGAGAGGATGAGGAGCATCTGCAGGTCAAGGGCCGAAAGCCTGAA	299
Db	241	GCAGCTGCTCAGGAGGAGAGGATGAGGAGCATCTGCAGGTCAAGGGCCGAAAGCCTGAA	300
QY	300	GCTATAGCCAGGAAACAGGTCACCCACACAGACTGGGTGTGAGTGTGAAGATGGTCTCGAT	359
Db	301	GCTATAGCCAGGAAACAGGTCACCCACACAGACTGGGTGTGAGTGTGAAGATGGTCTCGAT	360
QY	360	GGCGAGGAGATGACCCGCCAAATCCACAGGAGGTGAAACGCCTGAAGAAGGTGAAAG	419
Db	361	GGCGAGGAGGTGACCCGCCAAATCCACAGGAGGTGAAACGCCTGAAGAAGGTGAAAG	420
QY	420	CAATCACAGTGTAAAGAAAGACACGTTGAAATGATGCAGGCTGCCTATGTTGGAAT	479
Db	421	CAATCACAGTGTAAAGAAAGACACGTTGAAATGATGCAGGCTGCCTATGTTGGAAT	480
QY	480	TTGTTTCATTAAATTTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAAAGAAAAA	538
Db	481	TTGTTTCATTAAATTTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAAAGAAAAA	539
RESULT 5			
AAX90522			
ID	AAX90522 standard; cDNA; 532 BP.		
XX	AAX90522;		
AC	AAX90522;		
XX	AAX90522;		
DT	30-SEP-1999 (first entry)		
XX	GAGE-5 tumour rejection antigen clone nucleotide sequence.		
DE	Human leukocyte antigen; HLA-A29; tumour rejection antigen;		
KW	detection; therapy; pathological condition; cancer; CTL;		
KW	cytolytic T lymphocyte; GAGE; ss.		
XX	Homo sapiens.		
OS	Homo sapiens.		
XX	WO9937665-A1.		
EN	29-JUL-1999.		
PD	12-JAN-1999; 99WO-US00775.		
XX	23-JAN-1998; 98US-0012818.		
PR	(LUDW-) LUDWIG INST CANCER RES.		
PA	Boon-Falleur T, Debacker O, Van Den Eynde B, Van Der Bruggen P;		
XX	WPI; 1999-469111/39.		
XX	New isolated peptides which bind to HLA-A29 molecules, which are		
PT	tumour rejection antigens used for detection and therapy of		
PT	pathological conditions, e.g. cancer		

XX Example 13; Fig 4; 62pp; English.

XX

CC The present invention describes peptides which bind to human leukocyte

CC antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into

CC tumour rejection antigens. They can be used for detecting cytolytic T

CC lymphocytes (CTLs) in pathological conditions such as cancer and in

CC HLA-typing assays. Complexes of HLA-29 molecules and the peptides can

CC be used for stimulating CTLs in vivo. The present sequence represents

CC a GAGE tumour rejection antigen clone, from an example from the present

CC invention.

XX

XX Sequence 532 BP; 156 A; 111 C; 154 G; 111 T; 0 other;

Query Match 93.8%; Score 504.8; DB 20; Length 532;

Best Local Similarity 98.1%; Pred. No. 1.7e-136;

Matches 522; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

QY 10 AGCTGTGAGGCGAGTGTGTGTGTTCTCGCGTCGCGACTCTTTTCTCTACTGAGATT 69

DB 1 AGCTGTGAGGCGAGTGTGTGTGTTCTCGCGTCGCGACTCTTTTCTCTACTGAGATT 60

QY 70 CATCTGTGTGAATATGACTTCGCGAGGAGATGCGCC--TATCGGCCTAGACCAAGAC 126

DB 61 CATCTGTGTGAATATGACTTCGCGAGGAGATGCGCCATTATTTGGCCTAGACCAAGC 120

QY 127 GCTACGTAGAGCCTCCTGAAATGATTGGGCTTATCGGCGCCGAGCAGTTCAGTGATGAAG 186

DB 121 GCTATGTACAGCCTCCTGAAGTGATTGGGCTTATCGGCGCCGAGCAGTTCAGTGATGAAG 180

QY 187 TGGAAACACAGCAACACCTGAAGAGGGGAACAGCAACTCAACGTCAGGATCCTGAGCTG 246

DB 181 TGGAAACACAGCAACACCTGAAGAGGGGAACAGCAACTCAACGTCAGGATCCTGAGCTG 240

QY 247 CTCAGGAGGAGAGGATGAGGAGCATCTCGAGTCAAGGCCGGAAGCCTGAAGCTCAT 306

DB 241 CTCAGGAGGAGAGGATGAGGAGCATCTCGAGTCAAGGCCGGAAGCCTGAAGCTCAT 300

QY 307 GCCAGGAACAGGGTCAACCAACAGACTGGGTGTGAGTGTGAAGATGGTCTCGATGGGCAGG 366

DB 301 GCCAGGAACAGGGTCAACCAACAGACTGGGTGTGAGTGTGAAGATGGTCTCGATGGGCAGG 360

QY 367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCTGAAGAGAGGTGAAAAGCAATCAC 426

DB 361 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCTGAAGAGAGGTGAAAAGCAATCAC 420

QY 427 AGTGTTAAAGAGACACGTTGAAATGATCGAGGCTGCTCTATGTTCGAAATTTGTTCA 486

DB 421 AGTGTTAAAGAGAGGACGTTGAAATGATCGAGGCTGCTCTATGTTCGAAATTTGTTCA 480

QY 487 TTAATAATCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 538

DB 481 TTAATAATCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 532

RESULT 6

ABL66323

ID ABL66323 standard; DNA; 528 BP.

XX

AC ABL66323;

XX

DT 15-MAY-2002 (first entry)

XX

DE Lung cancer related gene sequence SEQ ID NO:4660.

XX

KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;

KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;

KW gene; ds.

XX

OS Homo sapiens.

XX

FN WO200194629-A2.

XX PD 31-DEC-1997. 97WO-US10850. 96US-0669161. (LUDW-) LUDWIG INST CANCER RES. Boon-Falleur T, Debacker O, Van Den Eynde B; WPI: 1998-076905/07. P-PSDB; AAW47603. Isolated nucleic acid encoding GAGE tumour rejection antigen precursor - processed by HLA-Cw6 molecules into peptides, useful to diagnose melanomas Example 13; Fig 4; 60pp; English. The present sequence encodes a GAGE-6 tumour rejection antigen precursor (TRAP). The protein is expressed in a number of tumours. In contrast the only normal tissue which expresses GAGE TRAP protein is testis. Several GAGE TRAPs have been identified (see AAV18717-21). The major difference between these proteins and GAGE-1 (AAV05540) is the absence of a stretch of 143 bases located at position 379 to 521 of the GAGE-1 TRAP sequence. The rest of the sequences show mismatches at various position, with the exception of GAGE-3 whose 5' end is totally different from the other GAGE cDNAs for the first 112 bases. This region of GAGE-3 cDNA contains a long repeat and a hairpin structure. The antigens can be used to diagnose melanomas, characterised by expression of a TRAP or presentation of a tumour rejection antigen. Antigens shed into blood or urine can be observed and then used to confirm a diagnosis of melanoma using cytolytic T cell clone proliferation methodologies. Other uses for the processed peptides, include HLA-typing assays for, e.g. skin graft or organ transplants. Sequence 539 BP; 157 A; 114 C; 157 G; 111 T; 0 other;

Query Match 93.0%; Score 500.6; DB 19; Length 539; Best Local Similarity 96.8%; Pred. No. 2.9e-135; Mismatches 14; Indels 3; Gaps 1;

QY 3 GCCAGGGAGCTGTGAGCAGTGTGTGTTCTCGCTCCGGACTCTTTTCCCTCTAC 62
DB 1 GCCAGGGAGCTGTGAGCAGTGTGTGTTCTCGCTCCGGACTCTTTTCCCTCTAC 60

QY 63 TCAGATTCACTGTGTGAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCTAGA 119
DB 61 TGAGATTCACTGTGTGAATATGAGTTGGCGAGGAAGATCGACCATTATTGGCCTAGA 120

QY 120 CCAAGACGCTAGTACGCTCTCTGAATGATTGGGCTATCGGCCCGAGCAGTTCAGT 179
DB 121 CCAAGGCACTATGTACAGCCTCTCTGAAGTGATTTGGGCTATGGGCCCGAGCAGTTCAGT 180

QY 180 GATGAGTGAACACAGCAACACCTCTGAAGAAGGGGACCACTCAACCTCAGGATCCT 239
DB 181 GATGAGTGAACACAGCCTCTCTGAAGAAGGGGACCACTCAACCTCAGGATCCT 240

QY 240 GCAGTGTCTCAGAGGAGAGGATGAGGAGCATCTGCGAGGTCAAGGCCGAGCCTGAA 299
DB 241 GCAGTGTCTCAGAGGAGAGGATGAGGAGCATCTGCGAGGTCAAGGCCGAGCCTGAA 300

QY 300 GCTCATAGCCAGCAAGGGTCAACCACAGACTGGGTGTGAGTGTGAAGATGGTCCGTGAT 359
DB 301 GCTCATAGCCAGCAAGGGTCAACCACAGACTGGGTGTGAGTGTGAAGATGGTCCGTGAT 360

QY 360 GGGCAGGAGATGGACCCGCAAAATCCAGAGGAGGTGAAACCGCTGGAAGGTGAAAG 419
DB 361 GGGCAGGAGATGGACCCGCAAAATCCAGAGGAGGTGAAACCGCTGGAAGGTGAAAG 420

QY 420 CAATCAGTGTGTTAAAGAAGACAGCTTGAATGATGCAGGCTCTCTATGTTGGAAT 479

DB 421 CAATCAGAGTGTAAAGAAGACACGCTGAATGATGAGGCTGCTCTATGTTGGAAT 480
QY 480 TTGTTCAATTAATAATCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAGAAAAAAA 538
DB 481 TTGTTCAATTAATAATCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAGAAAAAAA 539

RESULT 8
AAV18720
ID AAV18720 standard; cDNA; 532 BP.
XX AC AAV18720;
XX DT 30-JUL-1998 (first entry)
XX DE cDNA encoding GAGE-5 tumour rejection antigen precursor.
XX KW GAGE tumour rejection antigen precursor; TRAP; tumour;
XX KW diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
XX KW HLA-typing assay; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 75..429
XX FT /tag= a
XX FT /transl_except= (pos:189..191, aa:Ala)
XX FT /transl_except= (pos:192..194, aa:Thr)
XX FT
XX FN WO9749417-A1.
XX PD 31-DEC-1997.
XX PF 23-JUN-1997; 97WO-US10850.
XX PR 24-JUN-1996; 96US-0669161.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Boon-Falleur T, Debacker O, Van Den Eynde B;
XX WPI: 1998-076905/07.
XX P-PSDB; AAW47602.

Isolated nucleic acid encoding GAGE tumour rejection antigen precursor - processed by HLA-Cw6 molecules into peptides, useful to diagnose melanomas Example 13; Fig 4; 60pp; English. The present sequence encodes a GAGE-5 tumour rejection antigen precursor (TRAP). The protein is expressed in a number of tumours. In contrast the only normal tissue which expresses GAGE TRAP protein is testis. Several GAGE TRAPs have been identified (see AAV18717-21). The major difference between these proteins and GAGE-1 (AAV05540) is the absence of a stretch of 143 bases located at position 379 to 521 of the GAGE-1 TRAP sequence. The rest of the sequences show mismatches at various position, with the exception of GAGE-3 whose 5' end is totally different from the other GAGE cDNAs for the first 112 bases. This region of GAGE-3 cDNA contains a long repeat and a hairpin structure. The antigens can be used to diagnose melanomas, characterised by expression of a TRAP or presentation of a tumour rejection antigen. Antigens shed into blood or urine can be observed and then used to confirm a diagnosis of melanoma using cytolytic T cell clone proliferation methodologies. Other uses for the processed peptides, include HLA-typing assays for, e.g. skin graft or organ transplants. Sequence 532 BP; 154 A; 110 C; 156 G; 112 T; 0 other;

Query Match 92.3%; Score 496.8; DB 19; Length 532; Best Local Similarity 97.2%; Pred. No. 3.6e-134; Mismatches 12; Indels 3; Gaps 1;

DT	30-JUL-1998	(first entry)
XX	CDNA encoding GAGE-4 tumour rejection antigen precursor.	
XX	GAGE tumour rejection antigen precursor; TRAP; tumour;	
KW	diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;	
KW	HLA-typing assay; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	
DE	83..435	
XX	Location/Qualifiers	
FT	/*tag= a	
FT	/transl_except= (pos:197..199, aa:Ala)	
FT	/transl_except= (pos:200..202, aa:Thr)	
XX		
XX	W09749417-A1.	
PN		
XX	31-DEC-1997.	
XX		
XX	23-JUN-1997; 97WO-US10850.	
PP		
XX	24-JUN-1996; 96US-0669161.	
PR		
XX	(LUDW-) LUDWIG INST CANCER RES.	
PA		
XX	Boon-Falleur T, Debacker O, Van Den Eynde B;	
PI		
XX	WPI; 1998-076905/07.	
DR		
XX	P-PSDB; AAW47601.	
DR		
XX	Isolated nucleic acid encoding GAGE tumour rejection antigen	
PT	precursor - processed by HLA-Cw6 molecules into peptides, useful to	
PT	diagnose melanomas	
PT		
XX		
PS	Example 13; Fig 4; 60pp; English.	
XX		
CC	The present sequence encodes a GAGE-4 tumour rejection antigen	
CC	precursor (TRAP). The protein is expressed in a number of tumours. In	
CC	contrast the only normal tissue which expresses GAGE TRAP protein is	
CC	testis. Several GAGE TRAPs have been identified (see AAV19717-21). The	
CC	major difference between these proteins and GAGE-1 (AAV05540) is the	
CC	absence of a stretch of 143 bases located at position 379 to 521 of the	
CC	GAGE-1 TRAP sequence. The rest of the sequences show mismatches at	
CC	various position, with the exception of GAGE-3 whose 5' end is totally	
CC	different from the other GAGE cDNAs for the first 112 bases. This	
CC	region of GAGE-3 cDNA contains a long repeat and a hairpin structure.	
CC	The antigens can be used to diagnose melanomas, characterised by	
CC	expression of a TRAP or presentation of a tumour rejection antigen.	
CC	Antigens shed into blood or urine can be observed and then used to	
CC	confirm a diagnosis of melanoma using cytolytic T cell clone	
CC	proliferation methodologies. Other uses for the processed peptides,	
CC	include HLA-typing assays for, e.g. skin graft or organ transplants.	
XX		
XX	Sequence 541 BP; 158 A; 117 C; 154 G; 112 T; 0 other;	
XX		
XX	Query Match	
XX	Best Local Similarity 91.2%; Score 490.6; DB 19; Length 541;	
XX	Matches 523; Conservative 0; Mismatches 14; Indels 4; Gaps 2;	
QY	2 CGCCAGGAGCTGTGAGGCAAGTCTGTGTGCTCTCTCGCGTCCGGACTCTTTTCTCTA 61	
Db	1 CGCCACGGAGCTGTGAGGCAAGTCTGTGTGCTCTCTCGCGTCCCGACTCTTTTCTCTA 60	
QY	62 CTGAGATTCAATCTGTGTGAATAATGAGTTGGCGAGGAAGATCGACC---TATCGGCTAG 118	
Db	61 CTGAGATTCAATCTGTGTGAATAATGAGTTGGCGAGGAAGATCGACTATTATTGGCTAG 120	
QY	119 ACCRAGCGCTAGTAGAGCTCTGTAATGATTTGGCGCTATCGGCCCGAGAGTTTACG 178	
Db	121 ACCRAGGCGCTATGTACAGCTCTCTGAAATGATTTGGGCTATTCGCGCCCGAGAGTTTACG 180	
QY	179 TGATGAAGTGGAAACCAAGCAACACTGAAGAGGGGAACCAAGCAACTCAACGTGAGATCC 238	

SQ	Sequence 560 BP; 177 A; 119 C; 150 G; 114 T; 0 other;	PA	(LUDW-) LUDWIG INST CANCER RES.
XX	Query Match 79.8%; Score 429.2; DB 20; Length 560;	XX	Boon-Falleur T, Debacker O, Van Den Eynde B;
PI	Best Local Similarity 92.2%; Pred. No. 1.6e-114;	XX	WPI; 1998-076905/07.
XX	Matches 452; Conservative 0; Mismatches 38; Indels 0; Gaps 0;	DR	P-PSDB; AAW47600.
DR		DR	
XX		XX	Isolated nucleic acid encoding GAGE tumour rejection antigen precursor - processed by HLA-Cw6 molecules into peptides, useful to diagnose melanomas
PT		PT	Example 13; Fig 4; 60pp; English.
PT		PT	The present sequence encodes a GAGE-3 tumour rejection antigen precursor (TRAP). The protein is expressed in a number of tumours. In contrast the only normal tissue which expresses GAGE TRAP protein is testis. Several GAGE TRAPs have been identified (see AAV18717-21). The major difference between these proteins and GAGE-1 (AAV05540) is the absence of a stretch of 143 bases located at position 379 to 521 of the GAGE-1 TRAP sequence. The rest of the sequences show mismatches at various positions, with the exception of GAGE-3 whose 5' end is totally different from the other GAGE cDNAs for the first 112 bases. This region of GAGE-3 cDNA contains a long repeat and a hairpin structure. The antigens can be used to diagnose melanomas, characterised by expression of a TRAP or presentation of a tumour rejection antigen. Antigens shed into blood or urine can be observed and then used to confirm a diagnosis of melanoma using cytolytic T cell clone proliferation methodologies. Other uses for the processed peptides, include HLA-typing assays for, e.g. skin graft or organ transplants.
XX		XX	
SQ	Sequence 560 BP; 175 A; 119 C; 152 G; 114 T; 0 other;	SQ	
	Query Match 78.3%; Score 421.2; DB 19; Length 560;		Best Local Similarity 91.2%; Pred. No. 3.4e-112;
	Matches 447; Conservative 0; Mismatches 43; Indels 0; Gaps 0;		
Qy	49 TCTTTTCTCTACTGAGATTCATCTGTGTGAATATGAGTTGGCGAGGAAGATCGACCT 108	Qy	49 TCTTTTCTCTACTGAGATTCATCTGTGTGAATATGAGTTGGCGAGGAAGATCGACCT 108
Db	71 TCTTCTCGCCAACTCATATTTTCACACAGATGAATCTTCAGTAGAGGAAATCGACCTATT 130	Db	71 TCGTCTTCGCCAACTCATATTTTCACACAGATGAATCTTCAGTAGAGGAAATCGACCTATT 130
Qy	109 ATCGGCCTAGACCAAGACGCTACGTAGAGCCTCTGAAATGATGGGCTATCGGCGCG 168	Qy	109 ATCGGCCTAGACCAAGACGCTACGTAGAGCCTCTGAAATGATGGGCTATCGGCGCG 168
Db	131 ATTGSCCTAGACCAAGGCGCTATGTACAGCCTCTCTGAAGTGTATGGGCTATCGGCGCG 190	Db	131 ATTGSCCTAGACCAAGGCGCTATGTACAGCCTCTCTGAAGTGTATGGGCTATCGGCGCG 190
Qy	169 AGCAGTTCACTGATGAAGTGGAAACACAGCAACACCTGAAGAAGGGGAACACGCAACTCAAC 228	Qy	169 AGCAGTTCACTGATGAAGTGGAAACACAGCAACACCTGAAGAAGGGGAACACGCAACTCAAC 228
Db	191 AGCAGTTCACTGATGAAGTGGAAACACAGAGCCTCTCTGAAGAAGGGGAACACGCAACTCAAC 250	Db	191 AGCAGTTCACTGATGAAGTGGAAACACAGAGCCTCTCTGAAGAAGGGGAACACGCAACTCAAC 250
Qy	229 GTCAGGATCTCGAGCTGCTCAGGAGGAGAGGATGAGGAGCATCTGCAGGTCAAGGGC 288	Qy	229 GTCAGGATCTCGAGCTGCTCAGGAGGAGAGGATGAGGAGCATCTGCAGGTCAAGGGC 288
Db	251 GTCAGGATCTCGAGCTGCTCAGGAGGAGAGGATGAGGAGCATCTGCAGGTCAAGGGC 310	Db	251 GTCAGGATCTCGAGCTGCTCAGGAGGAGAGGATGAGGAGCATCTGCAGGTCAAGGGC 310
Qy	289 CGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCAACCCACAGACTGGGTGTGAGTGAAG 348	Qy	289 CGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCAACCCACAGACTGGGTGTGAGTGAAG 348
Db	311 CGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCAACCCACAGACTGGGTGTGAGTGAAG 370	Db	311 CGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCAACCCACAGACTGGGTGTGAGTGAAG 370
Qy	349 ATGCTCTGTATGGGCGAGGATGGACCCGCCAAATCCAGAGGAGGTGAACCCCTGAAG 408	Qy	349 ATGCTCTGTATGGGCGAGGATGGACCCGCCAAATCCAGAGGAGGTGAACCCCTGAAG 408
Db	371 ATGCTCTGTATGGGCGAGGATGGACCCGCCAAATCCAGAGGAGGTGAACCCCTGAAG 430	Db	371 ATGCTCTGTATGGGCGAGGATGGACCCGCCAAATCCAGAGGAGGTGAACCCCTGAAG 430
Qy	409 AAGGTGAAAGCAATCACAGTGTAAAAGAACACAGCTTGAATGATGCAGGCTGCTCCT 468	Qy	409 AAGGTGAAAGCAATCACAGTGTAAAAGAACACAGCTTGAATGATGCAGGCTGCTCCT 468
Db	431 AAGGTGAAAGCAATCACAGTGTAAAAGAACACAGCTTGAATGATGCAGGCTGCTCCT 490	Db	431 AAGGTGAAAGCAATCACAGTGTAAAAGAACACAGCTTGAATGATGCAGGCTGCTCCT 490
Qy	469 ATGTTGGAATTTGTTTCAATTAATCTCCCAATAAAGCTTACAGCCTTCTGCAAGAA 528	Qy	469 ATGTTGGAATTTGTTTCAATTAATCTCCCAATAAAGCTTACAGCCTTCTGCAAGAA 528
Db	491 ATGTTGGAATTTGTTTCAATTAATTTCTCCCAATAAAGCTTACAGCCTTCTGCAAGAA 550	Db	491 ATGTTGGAATTTGTTTCAATTAATTTCTCCCAATAAAGCTTACAGCCTTCTGCAAGAA 550
Qy	529 AAAAAAAAAA 538	Qy	529 AAAAAAAAAA 538
Db	551 AAAAAAAAAA 560	Db	551 AAAAAAAAAA 560
RESULT 12			
AAV18718			
ID	AAV18718 standard; cDNA; 560 BP.		
XX			
AC	AAV18718;		
XX			
DT	30-JUL-1998 (first entry)		
XX			
DE	cDNA encoding GAGE-3 tumour rejection antigen precursor.		
XX			
KW	GAGE tumour rejection antigen precursor; TRAP; tumour;		
KW	diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;		
KW	HLA-typing assay; ss.		
XX			
OS	Homo sapiens.		
XX			
FH	Location/Qualifiers		
FT	100..457		
FT	/*tag= a		
FT	/transl_except= (pos:217..219, aa:Ala)		
FT	/transl_except= (pos:220..222, aa:Thr)		
XX			
FN	W09749417-A1.		
XX			
PD	31-DEC-1997.		
XX			
PF	23-JUN-1997; 97WO-US10850.		
XX			
PR	24-JUN-1996; 96US-0669161.		
XX			

RESULT 13

AAZ97217
ID AAZ97217 standard; cDNA; 1024 BP.

XX

XX

AC AAZ97217;

XX

DT 18-APR-2000 (first entry)

XX

DE

XX

XX

XX

KW

KW

KW

KW

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

AAZ97217 standard; cDNA; 1024 BP.

AAZ97217;

18-APR-2000 (first entry)

Human prostate cancer differentially expressed gene #78.

Prostate cancer specific gene; cancer; tumour progression; diagnosis;
hyperproliferative cell growth; prostatic disorder; treatment;
metastatic prostate cancer; benign prostate hyperplasia; BPH; ss.

Homo sapiens.

WO9964594-A2.

16-DEC-1999.

10-JUN-1999; 99WO-US13181.

11-JUN-1998; 98US-0088877.

09-JUN-1999; 99US-0088877.

(CHIR) CHIRON CORP.

Astel JH, Carroll E, Endege WO, Ford DM, Monahan JE, Schlegel R;
Steinmann KE, Zhang J;
WPI; 2000-116541/10.

New isolated prostate cancer specific nucleic acids, used to develop
products for the diagnosis and treatment of cancer -

Claim 2; Page 107; 212pp; English.

This sequence represents a prostate cancer specific nucleic acid
sequence. The invention relates to a method for diagnosing cancer,
tumour progression, hyperproliferative cell growth or accompanying
biological and physical manifestations. The method involves contacting
the biological sample with a probe that comprises a sequence capable of
hybridising to any of the 339 nucleotide sequences given in the
specification (see AAZ97140-297478) and detecting duplex formation. The
products and methods of the invention can be used for the diagnosis,
prognosis, and treatment of cancer, tumour progression,
hyperproliferative cell growth, and accompanying physical and biological
manifestations. They can be used particularly for prostatic disorders
such as metastatic prostate cancer, localised prostate cancer, or benign
prostate hyperplasia (BPH).

Sequence 1024 BP; 272 A; 243 C; 275 G; 207 T; 27 other;
Query Match 74.2%; Score 399.4; DB 21; Length 1024;
Best Local Similarity 98.5%; Pred. No. 9.6e-106;
Matches 403; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 130 ACGTAGAGCCTCTGAAATGATGGGCTATGCGGCCGAGAGTTCAGTGAAGTGG 189
Db 100 AGGTACAGCCTCTGAAATGATGGGCTATGCGGCCGAGAGTTCAGTGAAGTGG 159
Qy 190 AACACAGACCTGAAAGAGGGGAACACAGCACTCAACGTCAGGATCTGCGAGTGGTC 249
Db 160 AACACAGACCTGAAAGAGGGGAACACAGCACTCAACGTCAGGATCTGCGAGTGGTC 219
Qy 250 AGGAGGAGAGATGAGGGAGCATCTGAGTCAAGGCCGAGCCTGAAGTCTATAGCC 309
Db 220 AGGAGGAGAGATGAGGGAGCATCTGAGTCAAGGCCGAGCCTGAAGTCTATAGCC 279
Qy 310 AGGAACAGAGGTACCCACAGACTGGGTGAGTGAAGATGGTCTGATGGCGAGGAGA 369
Db 280 AGGAACAGAGGTACCCACAGACTGGGTGAGTGAAGATGGTCTGATGGCGAGGAGA 339

Qy 370 TGGACCCGCCAAATCCAGAGGAGGTGAAAGCGCTGAAGAGGTGAAAGCAATCACAGT 429
Db 340 TGGACCCGCCAAATCCAGAGGAGGTGAAAGCGCTGAAGAGGTGAAAGCAATCACAGT 399
Qy 430 GTTAAAGAGACACGCTTGAATGATGACAGGCTGCTCTATGTTGGAATTTGTTTCATTA 489
Db 400 GTTAAAGAGAGGACGCTTGAATGATGAGCAGGCTGCTCTATGTTGGAATTTGTTTCATTA 459
Qy 490 AAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAGAAAAA 538
Db 460 AAATTCTCCCAATAAAGCTTTACAGCCTTCTGTAAGAAAAA 508

RESULT 14

AAZ97216/c
ID AAZ97216 standard; cDNA; 1024 BP.

XX

XX

AC AAZ97216;

XX

DT 18-APR-2000 (first entry)

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

AAZ97216 standard; cDNA; 1024 BP.

AAZ97216;

18-APR-2000 (first entry)

Human prostate cancer differentially expressed gene #77.

Prostate cancer specific gene; cancer; tumour progression; diagnosis;
hyperproliferative cell growth; prostatic disorder; treatment;
metastatic prostate cancer; benign prostate hyperplasia; BPH; ss.

Homo sapiens.

WO9964594-A2.

16-DEC-1999.

10-JUN-1999; 99WO-US13181.

11-JUN-1998; 98US-0088877.

09-JUN-1999; 99US-0088877.

(CHIR) CHIRON CORP.

Astel JH, Carroll E, Endege WO, Ford DM, Monahan JE, Schlegel R;
Steinmann KE, Zhang J;
WPI; 2000-116541/10.

New isolated prostate cancer specific nucleic acids, used to develop
products for the diagnosis and treatment of cancer -

Claim 2; Page 107; 212pp; English.

This sequence represents a prostate cancer specific nucleic acid
sequence. The invention relates to a method for diagnosing cancer,
tumour progression, hyperproliferative cell growth or accompanying
biological and physical manifestations. The method involves contacting
the biological sample with a probe that comprises a sequence capable of
hybridising to any of the 339 nucleotide sequences given in the
specification (see AAZ97140-297478) and detecting duplex formation. The
products and methods of the invention can be used for the diagnosis,
prognosis, and treatment of cancer, tumour progression,
hyperproliferative cell growth, and accompanying physical and biological
manifestations. They can be used particularly for prostatic disorders
such as metastatic prostate cancer, localised prostate cancer, or benign
prostate hyperplasia (BPH).

Sequence 1024 BP; 187 A; 290 C; 232 G; 270 T; 45 other;
Query Match 73.9%; Score 397.4; DB 21; Length 1024;
Best Local Similarity 98.0%; Pred. No. 3.6e-105;
Matches 401; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 130 ACGTAGAGCCTCTGAAATGATGGGCTATGCGGCCGAGAGTTCAGTGAAGTGG 189
Db 521 AGGTCAGCCTCTGAAATGATGGGCTATGCGGCCGAGAGTTCAGTGAAGCAGG 462

Qy	190	AACCAGCAACACTGAAGAGGGGAA	CCAGCAACTCAACGTCAGGATCCTGCAGTGCTC	249
Db	461	AACCAGCAACCTGAAGAGGGGAA	CCAGCAACTCAACGTCAGGATCCTGCAGTGCTC	402
Qy	250	AGGAGGGAGAGGATGAGGAGCATCT	CGAGGTCGAAGGCCGAAGGCTCATAGCC	309
Db	401	AGGAGGGAGAGGATGAGGAGCATNT	CGAGGTCGAAGGCCGAGGCTCATAGCC	342
Qy	310	AGGAACAGGGTCAACCCACAGACT	GGGTGTGAGTGTGAAGTGGTCCTGATGGCAGGAGA	369
Db	341	AGGAACAGGGTCACCCACAGACT	GGGTGTGAGTGTGAAGTGGTCAGGAGA	282
Qy	370	TGGACCCGCCAATCCAGAGGAGTG	AAAAAGCGCTGGAAGAGGTTGAAAAGCAATCACAGT	429
Db	281	TGGACCCGCCAATCCAGAGGAGTG	AAAAAGCGCTGGAAGAGGTTGAAAAGCAATCACAGT	222
Qy	430	GTAAAAAGAGCACGTTGAAATGAT	GCAGGCTGCTCCTATGTTGGAAATTTGTTTCATTA	489
Db	221	GTAAAAAGAGGCACGTTGAAATGAT	GCAGGCTGCTCCTATGTTGGAAATTTGTTTCATTA	162
Qy	490	AAATTCCTCCCAATAAAGCTTTAC	GAGCTTCTGCAAGAAAAA	538
Db	161	AAATTCCTCCCAATAAAGCTTTAC	GAGCTTCTGTAATAAAAAA	113

```

CC The TRA peptides can be used in the diagnosis of pathological
CC conditions, e.g. cancer, and for carrying out HLA-typing assays,
CC e.g. for skin grafts or organ transplants.
XX
SQ Sequence 646 BP; 188 A; 140 C; 174 G; 144 T; 0 other;

Query Match          69.6%; Score 374.4; DB 17; Length 646;
Best Local Similarity 99.7%; Pred. No. 1.4e-98;
Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 36 CTGCGGTCCGGACTCTTTTTCCTCTACTGAGATTTCATCTGTGAAATATGAGTTGGCGA 95
   |||||
Db 1 CTGCGGTCCGGACTCTTTTTCCTCTACTGAGATTTCATCTGTGAAATATGAGTTGGCGA 60

Qy 96 GGAAGATCGACCTTATCGGCCCTAGACCAAGACGCTACGTAGAGCCCTCTGAAATGATTGGG 155
   |||||
Db 61 GGAAGATCGACCTTATCGGCCCTAGACCAAGACGCTACGTAGAGCCCTCTGAAATGATTGGG 120

Qy 156 CCTATGCGGCCCGGAGCAGTTTCAGTGATGAAAGTGAACACGACCAACCTCTGAAAGAGGGGAA 215
   |||||
Db 121 CCTATGCGGCCCGGAGCAGTTTCAGTGATGAAAGTGAACACGACCAACCTCTGAAAGAGGGGAA 180

Qy 216 CCAGCAACTCAACGCTCAGGATCTCTGACGCTGCTCAGGAGGGAGAGGATCAGGGAGCATCT 275
   |||||
Db 181 CCAGCAACTCAACGCTCAGGATCTCTGACGCTGCTCAGGAGGGAGAGGATCAGGGAGCATCT 240

Qy 276 GCAGTCAAGGGCCGAAGCCTCGAAGCTCATAGCCAGGAAACAGGGTCAACCCACAGACTGGG 335
   |||||
Db 241 GCAGTCAAGGGCCGAAGCCTCGAAGCTCATAGCCAGGAAACAGGGTCAACCCACAGACTGGG 300

Qy 336 TGTGAGTGTGAGATGGTCTCTGATGGCCAGGAGATGGACCCGCCAAATCCAGAGGAGGTG 395
   |||||
Db 301 TGTGAGTGTGAGATGGTCTCTGATGGCCAGGAGATGGACCCGCCAAATCCAGAGGAGGTG 360

Qy 396 AAAACGCCCTGAAGAA 411
   |||||
Db 361 AAAACGCCCTGAAGAA 376

```


Result No.	Score	Query Match	Length	DB	ID	Description
1	119	18.4	1300	2	T03166	probable immediate
2	110	17.0	139	2	S70010	glutamate/proline-
3	110	17.0	221	2	S70009	glutamate/proline-
4	106	16.4	301	2	E29149	proline-rich prote
5	105.5	16.3	260	2	S22373	proline-rich prote
6	104	16.1	396	1	A59938	surface protein rh
7	103.5	16.0	171	2	A27307	proline-rich phosph
8	102.5	15.9	815	2	B30843	glutinin high mole
9	102.5	15.9	835	2	JN0689	glutenin, high-mol
10	102.5	15.9	830	2	S15720	glutenin high mole
11	102.5	15.9	848	2	S02262	glutenin high mole
12	102	15.8	652	2	PC4014	calcium binding 14
13	102	15.8	1560	2	T30282	calcium-binding pr
14	101.5	15.7	300	2	S19560	proline-rich prote
15	101	15.6	754	2	B97586	hypothetical prote
16	101	15.6	754	2	A23807	OmpA family protei
17	100.5	15.6	605	1	FW3YBA	beta-conglycinin a
18	100.5	15.6	847	2	F96531	hypothetical prote
19	100	15.5	1881	2	H95076	zinc metalloprotei
20	99.5	15.4	309	2	S10889	proline-rich prote
21	99.5	15.4	548	1	QFPEGL	neurofilament trip
22	99.5	15.4	605	2	S20007	beta-conglycinin a
23	99.5	15.4	1110	2	I51116	NP-180 - sea lamp
24	98	15.2	212	2	B36298	proline-rich prote
25	98	15.2	838	1	EEWTHW	glutenin, high mol
26	98	15.2	1606	2	T34073	paranemin - chicke
27	98	15.2	2649	2	T51023	hypothetical prote
28	97.5	15.1	166	2	B25372	salivary proline-r
29	97.5	15.1	3436	2	S53659	tequiment protein 6

Query Match 17.0%; Score 110; DB 2; Length 139;
Best Local Similarity 31.8%; Pred. No. 0.11;

RESULT 11
S02262
glutenin high molecular weight chain Dx5 - wheat

```
Query Match      15.8%; Score 102; DB 2; Length 1560;
Best Local Similarity 32.0%; Pred. NO. 3.8;
Matches 32; Conservative 13; Mismatches 35; Indels 20; Gaps 5;
```

Qy 34 EYEPATPEEGEPTAQFDPAAGGEDEGASAGGGPPEAHSQEQHQTGCET----- 88

RESULT 14

S19560
proline-rich protein MP4 - mouse
C/Species: Mus musculus (house mouse)
C/Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Aug-1999
C/Accession: S19560; S22570
R/Robertson, S.G.E.; Layfield, R.; Bannister, A.J.; McDonald, C.J.
Eur. J. Biochem. 202, 969-974, 1991
A>Title: Gene sequence of mouse B-type proline-rich protein MP4. Transcriptional

Query Match	15.7%;	Score 101.5;	DB 2;	Length 300;
Best Local Similarity	28.8%;	Pred. No. 0.91;		
Matches 38;	Conservative 10;	Mismatches 55;	Indels 29;	Gaps 7;

Qy	4	RGRSTYR-PRPRRYVEPPMIGP-MRPEQSDEVEPATPEEGEPATQKQDPAAAQGEDE	61
		:	
Dd	121	QGSSOORPFGNQGGPPPGGQQRRPFGNQGP--PPRGGQQRPPQCGNQGPPQ	178
		:	
Qy	62	GA-----SAGGPKPEAHSQ-----EQGHPTGCECEDGPD-----GOEMDPNP	101
		:	
Dd	179	GGHPHPPRFGNQGGPPGGGQQRRTPQGNQGGPPQGG-----GFQAPPRFGNQGGPPQ	233
		:	
Qy	102	EVKTPEEKEKQ	113
		:	
Dd	234	GPOGPPRTGNQO	245

RESULT 15
B97586
hypothetical protein AGR_C_3445 [imported] - Agrobacterium tumefaciens (strain C)
C/Species: Agrobacterium tumefaciens
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C/Accession: B97586
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.;
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Mar
Science 294, 2323-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacter

A;Reference number: A97359; PMID:11743194
A;Accession: B97586
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-754 <KUR>
A;Cross-references: GB:AE007859; PIDN:AAK87643.1; PID:G15156995; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_3445
A;Map position: circular chromosome

Query Match 15.6%; Score 101; DB 2; Length 754;
Best Local Similarity 27.7%; Pred. No. 2.3;
Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 5;
Qy 10 RPRPRRYVEPPMIGMPRPQFSDEVPATPEEGEPATQODPAAAQEGEDEGASAGQ-- 67
Db 182 RPKPKQAEQPAGEGEQRPERPRKEPAK---EPAAEQQ-PAARPENAEQPAKPREPA 236
Qy 68 -GPKPENHSQEQGHPTGCECEGPDGQEMD-----PPNPEEVK 105
Db 237 PGKKPQV---EKAPEQKAEPAEKPVPEKKPAPEPAAKEAPVPTPTAPRPPAPEAQF 293
Qy 106 TPEEGEKOSQ 115
Db 294 NPAPGRQPSSE 303

Search completed: April 7, 2003, 18:14:23
Job time : 46 secs

GenCore version 5.1.4 ps_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2003, 18:13:10 ; Search time 12 Seconds
(without alignments)
400.938 Million cell updates/sec

Title: US-09-782-745-27

Perfect score: 646

Sequence: 1 MSWRGRSTYRPRRYVEPP.....DPNPPEVKTPEGEKQSQ 116

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	646	100.0	116	1 GGE2_HUMAN	Q13066 homo sapien
2	637	98.6	116	1 GGE8_HUMAN	Q9ueu5 homo sapien
3	625.5	96.8	117	1 GGE7_HUMAN	O76087 homo sapien
4	616.5	95.4	117	1 GGE4_HUMAN	Q13068 homo sapien
5	612.5	94.8	117	1 GGE5_HUMAN	Q13069 homo sapien
6	608.5	94.2	117	1 GGE6_HUMAN	Q13070 homo sapien
7	598	92.6	138	1 GGE1_HUMAN	Q13065 homo sapien
8	589.5	91.3	118	1 GGE3_HUMAN	Q13067 homo sapien
9	266	41.2	146	1 GGB1_HUMAN	O75459 homo sapien
10	229.5	35.5	111	1 GGD3_HUMAN	Q96gt9 homo sapien
11	120	18.6	102	1 GGC1_HUMAN	O60829 homo sapien
12	110	17.0	221	1 NOL3_RAT	Q62881 rattus norv
13	100.5	15.6	605	1 GLCA_SOYEN	P13916 glycine max
14	99.5	15.4	548	1 NFL_FIG	P02547 sus scrofa
15	99	15.3	839	1 GLUT5_WHEAT	P10388 triticum ae
16	98	15.2	838	1 GLUT4_WHEAT	P08489 triticum ae
17	97.5	15.1	166	1 PRPC_HUMAN	P02810 homo sapien
18	97.5	15.1	220	1 NOL3_MOUSE	Q9dlx0 mus musculu
19	97.5	15.1	235	1 BAR6_CHITE	P08726 chironomus
20	97.5	15.1	541	1 NFL_RAT	P19527 rattus norv
21	97	15.0	261	1 PRP2_MOUSE	P05142 mus musculu
22	96.5	14.9	229	1 VBA_HPV08	P06425 human papil
23	96.5	14.9	617	1 ABP1_SACEX	P38479 saccharomyc
24	96	14.9	542	1 NFL_MOUSE	P08551 mus musculu
25	96	14.9	554	1 NFL_BOVIN	P02548 bos taurus
26	95.5	14.8	543	1 NFL_HUMAN	P07196 homo sapien
27	95.5	14.8	634	1 SCL_RAT	P24054 rattus norv
28	94.5	14.6	245	1 VBA_HPV05	P06924 human papil
29	94.5	14.6	247	1 PRB4_HUMAN	P10163 homo sapien
30	94	14.6	234	1 PRPM_HUMAN	P10161 homo sapien
31	93.5	14.5	276	1 PRPL_HUMAN	P10162 homo sapien
32	93.5	14.5	580	1 SYN3_HUMAN	O14594 homo sapien
33	93	14.4	174	1 PRPP_HUMAN	P81489 homo sapien

34	93	14.4	206	1	PRP3_RAT	P04474 rattus norv
35	92.5	14.3	183	1	PRPH_MESAU	P06680 mesocricetu
36	92	14.2	466	1	CNGA_RAT	P10354 rattus norv
37	92	14.2	466	1	SLX10_MOUSE	Q04888 mus musculu
38	92	14.2	533	1	SLX21_MOUSE	O09141 mus musculu
39	92	14.2	595	1	VIE1_MCMVS	P11210 murine cyto
40	91.5	14.2	245	1	VBA_HPV5B	P26550 human papil
41	91.5	14.2	579	1	SYN3_RAT	O70441 rattus norv
42	91.5	14.2	601	1	CORO_SCHPO	O13923 schizosacch
43	91.5	14.2	1099	1	NKX1_HUMAN	O60721 homo sapien
44	91.5	14.2	1298	1	ICP4_HSV11	P08392 herpes simp
45	91	14.1	417	1	VGLD_HSVBS	Q08100 bovine herp

ALIGNMENTS

RESULT 1
GGE2_HUMAN
ID GGE2_HUMAN STANDARD; PRT; 116 AA.
AC Q13066;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-2 protein (G antigen 2).
GN GAGE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=95378788; PubMed=7544395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.;
RT "A new family of genes coding for an antigen recognized by autologous
RT cytolytic T lymphocytes on a human melanoma.";
RL J. Exp. Med. 182:689-698(1995).
CC -!- FUNCTION: ANTIGEN, RECOGNIZED ON MELANOMA BY AUTOLOGOUS CYTOLYTIC
CC T LYMPHOCYTES.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19143; AAA82745.1; -;
DR Genew; HGNC:4099; GAGE2.
DR MIM; 604244; -;
KW Multigene family; Antigen.
SQ SEQUENCE 116 AA; 12786 MW; DD305D5CA29AF19A CRC64;

Query Match 100.0%; Score 646; DB 1; Length 116;
Best Local Similarity 100.0%; Pred. No. 3.5e-38;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	1	MSWRGRSTYRPRRYVEPP	PMIGMRPEQSDVEVPATPEGEPTATQRPAAQAGED	60
Db	1	MSWRGRSTYRPRRYVEPP	PMIGMRPEQSDVEVPATPEGEPTATQRPAAQAGED	60
OY	61	EGASAGQGPKEAHSQGHQHPQTGCECDGPDQEMDPNPPEVKTPPEGEKQSQ	116	
Db	61	EGASAGQGPKEAHSQGHQHPQTGCECDGPDQEMDPNPPEVKTPPEGEKQSQ	116	

RESULT 2

```

GGE8 HUMAN
ID _GGE8 HUMAN STANDARD; PRT; 116 AA.
AC Q9UEU5.
AT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GAGE-8 protein (G antigen 8).
GN GAGE8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1] _
RN SEQUENCE FROM N.A.
RP MEDLINE=99323388; PubMed=10397259;
RA De Backer O., Arden K.C., Boretti M., Vantomme V., De Smet C.,
RA Czechay S., Viars C.S., De Plaen E., Brasseur F., Chomez P.,
RA Van den Eynde B., Boon T., van der Bruggen P.;
RT "Characterization of the GAGE genes that are expressed in various
human cancers and in normal testis.";
PL Cancer Res. 59:3157-3165(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99323388; PubMed=10397259;
RA De Backer O., Arden K.C., Boretti M., Vantomme V., De Smet C.,
RA Czechay S., Viars C.S., De Plaen E., Brasseur F., Chomez P.,
RA Van den Eynde B., Boon T., van der Bruggen P.;
RT "Characterization of the GAGE genes that are expressed in various
human cancers and in normal testis.";
PL Cancer Res. 59:3157-3165(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -!- TISSUE SPECIFICITY: Not expressed in normal tissues, except in
testis, but expressed by a large proportion of tumors of various
histological origins.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF055473; AAC33676.1; -.
DR EMBL; BC018052; AAH18052.1; -.
DR Genew; HGNC:4106; GAGE8.
KW Multigene family.
SQ SEQUENCE 116 AA; 12764 MW; DD3052939E66F19A CRC64;

Query Match 98.68; Score 637; DB 1; Length 116;
Best Local Similarity 99.14; Pred. No. 1.4e-37;
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSWRGRSTYRPRRYVEPPPMRPFQSFDEVEPATPEGEPAQRQDPAAQEGED 60
DB 1 MSWRGRSTYRPRRYVEPPPMRPFQSFDEVEPATPEGEPAQRQDPAAQEGED 60

QY 61 EGASAGQGPKEAHSQBGHPQTGCCEGDPGQEMDPNPNPEVKTPPEGEKQSQ 116
DB 61 EGASAGQGPKEAHSQBGHPQTGCCEGDPGQEMDPNPNPEVKTPPEGEKQSQ 116

RESULT 3
GGE7 HUMAN
ID _GGE7 HUMAN STANDARD; PRT; 117 AA.
AC O76087;
AT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GAGE-7 protein (G antigen 7) (GAGE-7B) (GAGE-8) (AL4).
GN GAGE7B OR GAGE7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1] _
RN SEQUENCE FROM N.A.
RP

```

```

RX MEDLINE=98316329; PubMed=9651357;
RA Chen M.E., Lin S.-H., Chung L.W.K., Sikes R.A.;
RT "Isolation and characterization of PAGE-1 and GAGE-7: new genes
expressed in the LNCap prostate cancer progression model that share
homology with melanoma-associated antigens.";
RL J. Biol. Chem. 273:17618-17625(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99323388; PubMed=10397259;
RA De Backer O., Arden K.C., Boretti M., Vantomme V., De Smet C.,
RA Czechay S., Viars C.S., De Plaen E., Brasseur F., Chomez P.,
RA Van den Eynde B., Boon T., van der Bruggen P.;
RT "Characterization of the GAGE genes that are expressed in various
human cancers and in normal testis.";
PL Cancer Res. 59:3157-3165(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99323388; PubMed=10397259;
RA De Backer O., Arden K.C., Boretti M., Vantomme V., De Smet C.,
RA Czechay S., Viars C.S., De Plaen E., Brasseur F., Chomez P.,
RA Van den Eynde B., Boon T., van der Bruggen P.;
RT "Characterization of the GAGE genes that are expressed in various
human cancers and in normal testis.";
PL Cancer Res. 59:3157-3165(1999).
RN [2]
RP SEQUENCE FROM N.A.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN SOME PROSTATE CANCER TISSUES BUT
NOT IN NORMAL PROSTATE TISSUE.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF058988; AAC25989.1; -.
DR EMBL; AF055474; AAC33677.1; -.
DR EMBL; AF055475; AAC33678.1; -.
DR Genew; HGNC:4104; GAGE7.
DR Genew; HGNC:4105; GAGE7B.
DR MIM; 604132; -.
KW Multigene family.
SQ SEQUENCE 117 AA; 12978 MW; 365ED8D41361C7AD CRC64;

Query Match 96.8%; Score 625.5; DB 1; Length 117;
Best Local Similarity 97.4%; Pred. No. 8.8e-37;
Matches 114; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 MSWRGRSTYR-PRPRRYVEPPMIGMRPFQSFDEVEPATPEGEPAQRQDPAAQEGE 59
DB 1 MSWRGRSTYRPRRYVEPPPMRPFQSFDEVEPATPEGEPAQRQDPAAQEGE 60

QY 60 DEGASAGQGPKEAHSQBGHPQTGCCEGDPGQEMDPNPNPEVKTPPEGEKQSQ 116
DB 61 DEGASAGQGPKEAHSQBGHPQTGCCEGDPGQEMDPNPNPEVKTPPEGEKQSQ 117

RESULT 4
GGE4 HUMAN
ID _GGE4 HUMAN STANDARD; PRT; 117 AA.
AC Q13068;
AT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-4 protein (G antigen 4).
GN GAGE4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1] _
RN SEQUENCE FROM N.A.
RP TISSUE=Melanoma;
RX MEDLINE=95378788; PubMed=75443395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.;
RT "A new family of genes coding for an antigen recognized by autologous
cytolytic T lymphocytes on a human melanoma.";
RL J. Exp. Med. 182:689-698(1995).
RN [1]
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
NOT IN NORMAL TISSUES, EXCEPT TESTIS.

```

```
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19145; AAA82747.1; -.
DR Genew; HGNC:4101; GAGE4.
DR MIM; 604246; -.
KW Multigene family.
SQ SEQUENCE 117 AA; 12956 MW; 365ED71B2F9DC7AD CRC64;

  Query Match          95.4%; Score 616.5; DB 1; Length 117;
  Best Local Similarity 96.6%; Pred. No. 3.6e-36;
  Matches 113; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 1 MSWRGRSTYR-PRRRYVEPEMIGPMRPEQFSDEVEPATPEEGEPATQRDPAAAOEGE 59
Db 1 MSWRGRSTYYPRRRYVQPPPEMIGPMRPEQFSDEVEPATPEEGEPATQRDPAAAOEGE 60

Qy 60 DEGASAGQGPKEAHSQGHQPTGCECEDGPDGQEMDPPNPPEVKTPEEGEKQSQC 116
Db 61 DEGASAGQGPKEADSQGHQPTGCECEDGPDGQEMDPPNPPEVKTPEEGEKQSQC 117

RESULT 5
GGE5_HUMAN
ID -GGE5_HUMAN STANDARD; PRT; 117 AA.
AC Q13069;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE GAGE-5 protein (G antigen 5).
GN GAGE5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=95378788; PubMed=7544395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.;
RT "A new family of genes coding for an antigen recognized by autologous
RT cytolytic T lymphocytes on a human melanoma.";
RL J. Exp. Med. 182:689-698(1995).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19146; AAA82748.1; -.
DR EMBL; BC024914; AAH24914.1; -.
DR Genew; HGNC:4102; GAGE5.
DR MIM; 604247; -.
KW Multigene family.
```

```
SQ SEQUENCE 117 AA; 12924 MW; 234A865E2ECDD06 CRC64;

  Query Match          94.8%; Score 612.5; DB 1; Length 117;
  Best Local Similarity 95.7%; Pred. No. 6.8e-36;
  Matches 112; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 1 MSWRGRSTYR-PRRRYVEPEMIGPMRPEQFSDEVEPATPEEGEPATQRDPAAAOEGE 59
Db 1 MSWRGRSTYYPRRRYVQPPPEMIGPMRPEQFSDEVEPATPEEGEPATQRDPAAAOEGE 60

Qy 60 DEGASAGQGPKEAHSQGHQPTGCECEDGPDGQEMDPPNPPEVKTPEEGEKQSQC 116
Db 61 DEGASAGQGPKEADSQGHQPTGCECEDGPDGQEMDPPNPPEVKTPEEGEKQSQC 117

RESULT 6
GGE6_HUMAN
ID -GGE6_HUMAN STANDARD; PRT; 117 AA.
AC Q13070;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-6 protein (G antigen 6).
GN GAGE6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=95378788; PubMed=7544395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.;
RT "A new family of genes coding for an antigen recognized by autologous
RT cytolytic T lymphocytes on a human melanoma.";
RL J. Exp. Med. 182:689-698(1995).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19147; AAA82749.1; -.
DR Genew; HGNC:4103; GAGE6.
DR MIM; 604248; -.
KW Multigene family.
SQ SEQUENCE 117 AA; 12892 MW; 234A865E3FCCDD06 CRC64;

  Query Match          94.2%; Score 608.5; DB 1; Length 117;
  Best Local Similarity 94.9%; Pred. No. 1.3e-35;
  Matches 111; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Qy 1 MSWRGRSTYR-PRRRYVEPEMIGPMRPEQFSDEVEPATPEEGEPATQRDPAAAOEGE 59
Db 1 MSWRGRSTYYPRRRYVQPPPEMIGPMRPEQFSDEVEPATPEEGEPATQRDPAAAOEGE 60

Qy 60 DEGASAGQGPKEAHSQGHQPTGCECEDGPDGQEMDPPNPPEVKTPEEGEKQSQC 116
Db 61 DEGASAGQGPKEADSQGHQPTGCECEDGPDGQEMDPPNPPEVKTPEEGEKQSQC 117

RESULT 7
GGE7_HUMAN
ID -GGE7_HUMAN STANDARD; PRT; 138 AA.
AC Q13065;
DT 15-JUL-1998 (Rel. 36, Created)
```

DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-1 protein (G antigen 1) (M22-F antigen).
GN GAGE1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=95378788; PubMed=7544395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.; family of genes coding for an antigen recognized by autologous
RT cytolytic T lymphocytes on a human melanoma.";
RL J. Exp. Med. 182:689-698(1995).
CC -!- FUNCTION: ANTIGEN, RECOGNIZED ON MELANOMA BY AUTOLOGOUS CYTOLYTIC
CC T LYMPHOCYTES. COMPLETELY SILENT IN NORMAL ADULT TISSUES, EXCEPT
CC TESTIS
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19142; AAA82744.1; -;
DR Genew; HGNC:4098; GAGE1.
DR MIM; 604243; -;
KW Multigene family; Antigen.
SQ SEQUENCE 138 AA; 15418 MW; 37B8F3909EC4B3B2 CRC64;
Query Match 92.6%; Score 598; DB 1; Length 138;
Best Local Similarity 99.1%; Pred. No. 7.6e-35;
Matches 108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSWRGSTRPRRRVPEPMIGMPRPFQSFDSVEPATPEEGEPATQRPAAQEGED 60
Db |||||
QY 1 MSWRGSTRPRRRVPEPMIGMPRPFQSFDSVEPATPEEGEPATQRPAAQEGED 60
Db |||||
QY 61 EGASAGQGPKEAHSQEQGHQPGTGCEDGPDGQMDPPNPPEVKTPEE 109
Db |||||
QY 61 EGASAGQGPKEAHSQEQGHQPGTGCEDGPDGQMDPPNPPEVKTPEE 109
Db |||||
RESULT 8
GGE3 HUMAN
ID GGE3 HUMAN STANDARD; PRT; 118 AA.
AC Q13067;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-3 protein (G antigen 3).
GN GAGE3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=95378788; PubMed=7544395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.;
RA "A new family of genes coding for an antigen recognized by autologous
RT cytolytic T lymphocytes on a human melanoma.";
RL J. Exp. Med. 182:689-698(1995).

CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19144; AAA82745.1; -;
DR Genew; HGNC:4100; GAGE3.
DR MIM; 604245; -;
KW Multigene family.
SQ SEQUENCE 118 AA; 12937 MW; D97ERBB19E735103 CRC64;
Query Match 91.3%; Score 589.5; DB 1; Length 118;
Best Local Similarity 94.7%; Pred. No. 2.5e-34;
Matches 106; Conservative 3; Mismatches 2; Indels 1; Gaps 1;
QY 4 RGRSTYR-PRPRRYVEPMIGMPRPFQSFDSVEPATPEEGEPATQRPAAQEGED 62
Db |||||
QY 5 RKGSTYWPRRYVQPPVIGMPRPFQSFDSVEPATPEEGEPATQRPAAQEGED 64
Db |||||
QY 63 ASAGQGPKEAHSQEQGHQPGTGCEDGPDGQMDPPNPPEVKTPEEGEKQSQ 116
Db |||||
QY 65 ASAGQGPKEAHSQEQGHQPGTGCEDGPDGQMDPPNPPEVKTPEEGEKQSQ 118
Db |||||
RESULT 9
GGE1 HUMAN
ID GGE1 HUMAN STANDARD; PRT; 146 AA.
AC Q75459; Q9BSS7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE G antigen family B 1 protein (Prostate-associated gene protein 1)
DE (PAGE-1) (GAGE-9) (AL5).
GN GAGEB1 OR PAGE1 OR GAGE9.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98316329; PubMed=9651357;
RA Chen M.E., Lin S.-H., Chung L.W.K., Sikes R.A.;
RT "Isolation and characterization of PAGE-1 and GAGE-7: new genes
RT expressed in the LNCap prostate cancer progression model that share
RT homology with melanoma-associated antigens.";
RL J. Biol. Chem. 273:17618-17625(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: UNKNOWN.
CC -!- TISSUE SPECIFICITY: ISOLATED FROM PROSTATE CANCER CELL LINES;
CC EXPRESSION ASSOCIATED WITH PROGRESSION TO ANDROGEN INSENSITIVE
CC PHENOTYPE. EXPRESSED IN NORMAL TESTIS AND AT LOWER LEVEL IN NORMAL
CC PLACENTA.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL; AF058989; AAC25990.1; -.
DR EMBL; BC004861; AAH04861.1; -.
DR Genew; HGNC:4107; GAGEB1.
DR MIM; 300288; -.
KW Multigene family.
FT CONFLICT 75 P -> L (IN REF. 2).
SQ SEQUENCE 146 AA; 16134 MW; 91ABA1E3B498DCA2 CRC64;

Query Match
Best Local Similarity 41.2%; Score 266; DB 1; Length 146;
Matches 66; Conservative 7; Mismatches 30; Indels 54; Gaps 3;

QY 1 MSWRGRSTYRPRRYVEPPMIGMRPEQFSDEVEPA--TPEGEPTATQRQDPAQAQEG 60
DB 1 MGFLRLRIYRRFIYVE-----SSEESDE---QPDEVESPTQSDSTPAERED 48
QY 61 EGASAGO-----GPKPEAHSQEQ 78
DB 49 EGASAGQPEADSQELVQPKTGCPEGDPDTKRVCLRNEEQMKLPAGGFEADSQEQ 108
QY 79 GHPQTGCEBDDPGQEMDPNPPNPEVKTPPEGEKQSQ 115
DB 109 VHPKTGCEGDDPVQELGLFNPEVKTPPEDEGQSQ 145

RESULT 10
GGD3_HUMAN
ID GGD3_HUMAN STANDARD; PRT; 111 AA.
AC Q96GT9;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE G antigen family D 3 protein (XAGE-2).
GN GAGED3 OR XAGE3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zandman A.J.W., van Kraats A.A., Weidle U.H., Ruiter D.J.,
RA Van Muijen G.N.P.;
RT "Expression profile and alignment of the XAGE family of cancer/testis
associated genes.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBSJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RC TISSUE=Placenta;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBSJ databases.
CC -!- FUNCTION: UNKNOWN.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AJ318891; CAC88125.1; -.
DR EMBL; AJ318892; CAC88125.1; JOINED.
DR EMBL; AJ318880; CAC83007.1; -.
DR EMBL; BC009232; AAH09232.1; -.
DR Genew; HGNC:4112; GAGED3.
KW Multigene family.
SQ SEQUENCE 111 AA; 12354 MW; 38C93332C5BA0E14 CRC64;

Query Match
Best Local Similarity 35.5%; Score 229.5; DB 1; Length 111;
Matches 54; Conservative 15; Mismatches 38; Indels 11; Gaps 3;

```

```

QY 1 MSWRGRSTYRPRRYVEPPMIGMRPEQFSDEVEPA--TPEGEPTATQRQDPAQAQEG 58
DB 1 MSWRGRSTYRPRRSLQPELGAM-----LEPTDEEPKEKPPTKSRNTPDQKR 52
QY 59 E-DEGASAGQPKPEAHSQEQHQPTGCEBDDPGQEMDPNPPNPEVKTPPEGEKQSQ 115
DB 53 EDDQGAABIQVPDLEADLQELCQTKTGDCGEGTDVKGLPKAHEHFKMPEAGEGKQSQ 110

RESULT 11
GGC1_HUMAN
ID GGC1_HUMAN STANDARD; PRT; 102 AA.
AC O60829;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE G antigen family C 1 protein (Prostate-associated gene protein 4)
DE (PAGE-4) (PAGE-1) (JM27) (GAGE-9).
GN GAGEC1 OR PAGE4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98393718; PubMed=9724777;
RA Brinkmann U., Vasmatazis G., Lee B., Yerushalmi N., Essand M.,
RA Pastan I.;
RT "PAGE-1, an X chromosome-linked GAGE-like gene that is expressed in
normal and neoplastic prostate, testis, and uterus.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10757-10762(1998).
[2]
RP SEQUENCE FROM N.A.
RA Strom T.M., Nyakatura G., Hellebrand H., Drescher B., Rosenthal A.,
RA Meindl A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBSJ databases.
[3]
RP SEQUENCE FROM N.A.
RA TISSUE=Placenta;
RC TISSUE=Placenta;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBSJ databases.
CC -!- FUNCTION: UNKNOWN.
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN NORMAL MALE AND
FEMALE REPRODUCTIVE TISSUES, PROSTATE, TESTIS, FALLOPIAN TUBE,
UTERUS, AND PLACENTA, AS WELL AS IN PROSTATE CANCER, TESTICULAR
CANCER, AND UTERINE CANCER.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AF275258; AAF8037.1; -.
DR EMBL; AJ005894; CAA06751.1; -.
DR EMBL; AF230380; AAF62541.1; -.
DR EMBL; BC010897; AAH10897.1; -.
DR Genew; HGNC:4108; GAGEC1.
DR MIM; 300287; -.
KW Multigene family.
SQ SEQUENCE 102 AA; 11153 MW; CE5D07AFBF73301B CRC64;

Query Match
Best Local Similarity 18.6%; Score 120; DB 1; Length 102;
Matches 39; Conservative 10; Mismatches 52; Indels 12; Gaps 3;

QY 1 MSWRGRSTYRPRRYVEPPMIGMRPEQFSDEVEPATPEGEPTATQRQDPAQAQEGED 60
DB 1 MSARVRSRSGRGDQ--EAPDVVA-----FVAPGSSQOEPTNDQIEFGQ--ER 48

```


OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=85154583; PubMed=3920075;
 RA Geisler N., Plessmann U., Weber K.;
 RT "The complete amino acid sequence of the major mammalian
 RL neurofilament protein (NF-L).";
 RL FEBS Lett. 182:475-478 (1985).
 RP [2]
 RN SEQUENCE OF 1-82 AND 278-548.
 RA Geisler N., Kaufmann E., Fischer S., Plessmann U., Weber K.;
 RT "Neurofilament architecture combines structural principles of
 FT intermediate filaments with carboxy-terminal extensions increasing
 FT in size between triplet proteins.";
 RL EMO J. 2:1295-1302 (1983).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC -1- DOMAIN: THE EXTRA MASS AND HIGH CHARGE DENSITY THAT DISTINGUISH
 CC THE NEUROFILAMENT PROTEINS FROM ALL OTHER INTERMEDIATE FILAMENT
 CC PROTEINS ARE DUE TO THE TAILPIECE EXTENSIONS. THIS REGION MAY FORM
 CC A CHARGED SCAFFOLDING STRUCTURE SUITABLE FOR INTERACTION WITH
 CC OTHER NEURONAL COMPONENTS OR IONS.
 CC -1- MISCELLANEOUS: NF-L IS THE MOST ABUNDANT OF THE THREE
 CC NEUROFILAMENT PROTEINS AND, AS THE OTHER NONEPITHELIAL
 CC INTERMEDIATE FILAMENT PROTEINS, IT CAN FORM HOMOPOLYMERIC 10-NM
 CC FILAMENTS.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 DR PIR; A02963; QPQGL.
 DR InterPro; IPR001664; IF.
 DR Pfam; PF00038; filament; 1.
 DR PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neurone; Glycoprotein.
 FT DOMAIN 1 91
 FT HEAD.
 FT DOMAIN 92 395
 FT ROD.
 FT DOMAIN 396 548
 FT TAIL.
 FT DOMAIN 92 123
 FT COIL 1A.
 FT DOMAIN 124 136
 FT LINKER 1.
 FT DOMAIN 137 232
 FT COIL 1B.
 FT DOMAIN 233 251
 FT LINKER 12.
 FT DOMAIN 252 270
 FT COIL 2A.
 FT DOMAIN 271 279
 FT LINKER 2.
 FT DOMAIN 280 395
 FT COIL 2B.
 FT DOMAIN 396 442
 FT TAIL, SUBDOMAIN A. (ACIDIC).
 FT DOMAIN 443 548
 FT TAIL, SUBDOMAIN B. (ACIDIC).
 FT CARBOHYD 20 20
 FT CARBOHYD 26 26
 FT CARBOHYD 26 26
 FT SITE 380 390
 FT MONOCLONAL ANTIBODY).
 FT OR K.
 FT UNSURE 322 322
 FT SEQUENCE 548 AA; 61940 MW; 83044813637ACT739 CRC64;
 Query Match 15.4%; Score 99.5; DB 1; Length 548;
 Best Local Similarity 26.2%; Pred. No. 2.2;
 Matches 34; Conservative 17; Mismatches 54; Indels 25; Gaps 4;
 QY 5 GRSTV-----RPRRY-----VEPEMIGMPREQFSDEVEPATPERGE 44
 DB 419 GRSYGGLOTSSYLMSTRSFPYSYTHVQEEQIEVEETIEAFAKAEAKDE-PPSEGEAE 477
 QY 45 PATQRQDPAQAQEGDEGASAGQPKPEAHSQEQGHQPTQCECEDGPDGQMDPPNPREV 104
 DB 478 EGKEKEAE 104
 QY 105 KTEPEGEKQS 114
 DB 534 KQEGAGEBEQA 543
 RESULT 15
 GLT5_WHEAT
 ID GLT5_WHEAT STANDARD; PRT; 839 AA.
 AC P10388;
 DT 01-MAR-1989 (Rel. 10, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Glutinin, high molecular weight subunit DX5 precursor.
 GN GLU-1D-1D OR GLU-D1-1B.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=cv. Cheyenne;
 RX MEDLINE=89098419; PubMed=2563152;
 RA Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R.,
 RA Malpica-Romero J.M.;
 RT "Nucleotide sequences of the two high-molecular-weight glutenin genes
 RT from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv
 RT Cheyenne.";
 RL Nucleic Acids Res. 17:461-462 (1989).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RC STRAIN=cv. Cheyenne;
 RA Anderson O.D.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
 CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
 CC VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
 CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
 CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
 CC GROUP 1 CHROMOSOMES OF WHEAT.
 CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
 CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQPGQG AND
 CC GQPGQGQGGQGYPTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.16b-sib.ch/announcement/>
 CC or send an email to license@16b-sib.ch).
 CC -----
 DR EMBL; X12928; CAA31395.1; -.
 DR PIR; S02262; S02262.
 DR InterPro; IPR001419; Glutinin.
 DR Pfam; PF03157; Glutinin.bmw; 1.
 DR PRINTS; PR00210; GLUTENIN.
 KW Seed storage protein; Repeat; Multigene family; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 839 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
 FT DX5.
 FT DOMAIN 131 801 REPEATS.
 FT SEQUENCE 839 AA; 89359 MW; 0F14E1106D552643 CRC64;
 Query Match 15.3%; Score 99; DB 1; Length 839;
 Best Local Similarity 23.5%; Pred. No. 3.5;
 Matches 31; Conservative 22; Mismatches 57; Indels 22; Gaps 5;
 QY 4 RGRSTYR-----RPRRYVEPPE-----MTGPMRPEQS----DEVEPATPERGE 44
 DB 491 QGQPGYPTSPQSGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG 550
 QY 45 PATQRQDPAQAQEGED-EGASAGQPKPEAHSQEQGHQPTQCECEDGPDGQMDPPNPREE 103
 DB 551 QPGQLQPAQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGGQGG 608
 QY 104 VKTPEGEKQS 115
 DB 609 GQOPAQGGQGGQ 620
 Search completed: April 7, 2003, 18:13:32
 Job time : 14 secs

us-09-782-745-27.rsp

Wed Apr 9 18:05:30 2003

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2003, 18:13:11 ; Search time 33 Seconds
(without alignments)
724.287 Million cell updates/sec

Title: US-09-782-745-27
Perfect score: 646
Sequence: 1 MSWRGRSTRPRRYVEPP.....DPPNPEVKTPEGEKQSQ 116

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 21:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phase:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_rvirus:.*
16: sp_bacteriap:.*
17: sp_archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	637	98.6	116	Q9UEUS	Q9ueus homo sapien
2	271.5	42.0	111	Q8WTF9	Q8wtf9 homo sapien
3	266	41.2	146	Q9BS87	Q9bs87 homo sapien
4	262.5	40.6	111	Q8WYS9	Q8wys9 homo sapien
5	229.5	35.5	111	Q96GT9	Q96gt9 homo sapien
6	199	30.8	108	Q8WMM1	Q8wmm1 homo sapien
7	124.5	19.3	69	Q8WMM0	Q8wmm0 homo sapien
8	119	18.4	1300	O36421	O36421 alcelaphine
9	117.5	18.2	128	O18563	O18563 strongyloid
10	116.5	18.0	712	Q8RWX5	Q8rwx5 arabidopsis
11	116.5	18.0	956	Q9LJ64	Q9lj64 arabidopsis
12	115.5	17.9	144	O18564	O18564 strongyloid
13	115.5	17.9	795	Q8S3W0	Q8s3w0 aegilops ma
14	112	17.3	504	Q91X93	Q91x93 mus musculu
15	111	17.2	1054	Q9JW99	Q9jw99 mus musculu
16	110	17.0	139	Q62882	Q62882 rattus norv

17	109	16.9	992	10	Q8S3V8	Q8s3v8 aegilops um
18	107.5	16.6	110	4	Q96GU1	Q96gu1 homo sapien
19	107.5	16.6	130	4	Q8WML9	Q8wml9 homo sapien
20	106	16.4	301	11	O62105	O62105 mus musculu
21	105.5	16.3	224	12	O73450	O73450 human papil
22	105.5	16.3	260	11	Q64306	Q64306 mus musculu
23	105.5	16.3	836	10	Q8S3T3	Q8s3t3 aegilops ta
24	105.5	16.3	910	11	Q9JLE8	Q9jle8 mus musculu
25	105	16.3	737	10	Q941L4	Q941l4 secale cere
26	104.5	16.2	912	11	Q9JLE7	Q9jle7 mus musculu
27	104.5	16.2	1343	12	O06635	O06635 bovine herp
28	104	16.1	396	5	O04151	O04151 toxoplasma
29	104	16.1	926	3	O13305	O13305 pneumocysti
30	103	15.9	836	10	Q8S3T2	Q8s3t2 aegilops ta
31	102.5	15.9	815	10	Q41553	Q41553 triticum ae
32	102.5	15.9	830	10	Q03872	Q03872 triticum ae
33	102.5	15.9	999	11	Q9JKR6	Q9jkr6 mus musculu
34	102	15.8	480	5	Q27033	Q27033 theileria p
35	102	15.8	614	5	O94674	O94674 plasmodium
36	102	15.8	652	11	Q64139	Q64139 mus sp. cbp
37	102	15.8	1560	5	Q26644	Q26644 strongyloce
38	102	15.8	2047	4	O15019	O15019 homo sapien
39	101.5	15.7	300	11	Q61888	Q61888 mus musculu
40	101.5	15.7	543	5	Q27043	Q27043 theileria p
41	101	15.6	754	16	Q8U884	Q8u884 agrobacteri
42	101	15.6	897	3	Q96VJ1	Q96vj1 pneumocysti
43	100.5	15.6	847	10	Q9X1B6	Q9x1b6 arabidopsis
44	100.5	15.6	1927	2	Q54875	Q54875 streptococc
45	100	15.5	1881	16	Q9L7Q2	Q9l7q2 streptococc

ALIGNMENTS

RESULT 1	Q9UEUS	PRELIMINARY;	PRT;	116 AA.
ID	Q9UEUS			
AC	Q9UEUS;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	GAGE-8 (G antigen 8)			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=9323388; PubMed=10397259;			
RA	De Backer O., Arden K.C., Boretta M., Vantomme V., De Smet C.,			
RA	Czekay S., Viars C.S., De Plaen E., Brasseur F., Chomez P.,			
RA	Van den Bynde B., Boon T., van der Bruggen P.;			
RT	"Characterization of the GAGE genes that are expressed in various			
RT	human cancers and in normal testis."			
RL	Cancer Res. 59:3157-3165(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRIN;			
RA	Strausberg R.;			
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF055473; AAC33676.1; -			
DR	EMBL; BC018052; AAH18052.1; -			
DR	PRINTS; PR01574; TUBBYPROTEIN.			
SQ	SEQUENCE 116 AA; 12764 MW; DD3052939E66F19A CRC64;			

Query Match	98.6%;	Score 637;	DB 4;	Length 116;
Best Local Similarity	99.1%;	Pred. No. 1.2e-47;		
Matches 115;	Conservative	0;	Mismatches 1;	Indels 0;
QY	1 MSWRGRSTRPRRYVEPPPMRPFQSFDEVEPATPEGEPTATQRQDPAAQEGED 60			
Db	1 MSWRGRSTRPRRYVEPPPMRPFQSFDEVEPATPEGEPTATQRQDPAAQEGED 60			

QY 61 EGASAGQGPKEAHSQGHQHPQTGCECEDGPDGQEMDPPNPPEVKTPEEGEKQSQ 116
 |||||
 Db 61 EGASAGQGPKEADSQGHQHPQTGCECEDGPDGQEMDPPNPPEVKTPEEGEKQSQ 116
 |||||

RESULT 2

Q9WTP9
 ID Q8WTP9 PRELIMINARY; PRT; 111 AA.
 AC Q8WTP9;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE XAGE-3 protein.
 GN XAGE-3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zendaan A.J.W., van Kraats A.A., Weidle U.H., Ruiter D.J.,
 RA Van Muijen G.N.P.;
 RT "Expression profile and alignment of the XAGE family of cancer/testis
 RT associated genes."
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=EWING SARCOMA;
 RA Zendaan A.J.W., van Kraats A.A., Weidle U.H., Ruiter D.R.,
 RA Van Muijen G.N.P.;
 RT "Expression profile of members of the XAGE cancer/testis antigen
 RT family."
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ318893; CAC83014.1; -.
 DR EMBL; AJ318881; CAC83008.1; -.
 SQ SEQUENCE 111 AA; 12302 MW; F4A3261B6F7ACFSB CRC64;

Query Match 42.0%; Score 271.5; DB 4; Length 111;
 Best Local Similarity 50.0%; Pred. No. 2.5e-16;
 Matches 59; Conservative 13; Mismatches 35; Indels 11; Gaps 3;

QY 1 MSWRGSTRPRRRYVEPEMIGMPRPEQFSDEVEPA--TPEEGEPATQODPAAQOE- 57
 |||||
 Db 1 MWRRGSTRPRRRSRVPPPELIGPM-----LEPDEBPQQEPPPTESRDPAPQOE 52
 |||||

RESULT 3

Q9BSS7
 ID Q9BSS7 PRELIMINARY; PRT; 146 AA.
 AC Q9BSS7;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE G antigen, family B, 1 (prostate associated).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SKIN;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC004861; AAH04861.1; -.
 SQ SEQUENCE 146 AA; 16150 MW; B6C7BA94D898DCE2 CRC64;

Query Match 41.2%; Score 266; DB 4; Length 146;
 Best Local Similarity 42.0%; Pred. No. 9.9e-16;
 Matches 66; Conservative 10; Mismatches 30; Indels 54; Gaps 3;

QY 1 MSWRGSTRPRRRYVEPEMIGMPRPEQFSDEVEPATPEEGEPATQODPAAQOE 60
 |||||
 Db 1 MGFLRLIYRRPMIYVE-----SSESSDE---QPDEVESPTQSDSTPAERED 48
 |||||

RESULT 4

Q8WYS9
 ID Q8WYS9 PRELIMINARY; PRT; 111 AA.
 AC Q8WYS9;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical 12.3 kDa protein.
 GN PP9012.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Zhou X.M., Zhang P.P., Jiang H.Q., Huang Y., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF318372; AAL55879.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 111 AA; 12342 MW; BD4A7E6788458B2 CRC64;

Query Match 40.6%; Score 262.5; DB 4; Length 111;
 Best Local Similarity 49.2%; Pred. No. 1.5e-15;
 Matches 58; Conservative 13; Mismatches 36; Indels 11; Gaps 3;

QY 1 MSWRGSTRPRRRYVEPEMIGMPRPEQFSDEVEPA--TPEEGEPATQODPAAQOE- 57
 |||||
 Db 1 MWRRGSTRPRRRSRVPPPELIGPM-----LEPDEBPQQEPPPTESRDPAPQOE 52
 |||||

RESULT 5

Q96GT9
 ID Q96GT9 PRELIMINARY; PRT; 111 AA.
 AC Q96GT9;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Similar to G antigen 8 (XAGE-2 protein).
 GN XAGE-2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Strausberg R.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Zendaan A.J.W., van Kraats A.A., Weidle U.H., Ruiter D.J.,
 RA Van Muijen G.N.P.;

```
RT "Expression profile and alignment of the XAGE family of cancer/testis
RT associated genes.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
[3]
RC TISSUE=EWING SARCOMA;
RA Zandman A.J.W., van Kraats A.A., Weidle U.H., Ruiter D.R.,
RA Van Muijen G.N.P.;
RT "Expression profile of members of the XAGE cancer/testis antigen
RT family.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009232; AAH0932.1; -.
DR EMBL; AJ318891; CAC88125.1; -.
DR EMBL; AJ318892; CAC88125.1; JOINED.
DR EMBL; AJ318880; CAC83007.1; -.
SQ SEQUENCE 111 AA; 12354 MW; 38C93332C5BA0E14 CRC64;

Query Match 35.5%; Score 229.5; DB 4; Length 111;
Best Local Similarity 45.8%; Pred. No. 1e-12;
Matches 54; Conservative 15; Mismatches 38; Indels 11; Gaps 3;

QY 1 MSNRGRSTYPRPRRYVEPPMIGMRPEQFSDEVEPA--TPEGEPATORQDPAAQAQEG 58
Db 1 MSNRGRSTYPRPRRSQPPPLIGAM-----LEPTDEEPKEKPPTKSRNPTPDQKR 52
QY 59 E-DEGASAGQPKPEAHSQEQGHPTQGCEDGPDGQEMDPNPEVKTTPERGEKQSQ 115
Db 53 EDQGAIEIQVPLEADLQELCQTKTGCGEGTDVKGKILPKAEHFKMPEAGEGKQSQ 110

RESULT 6
Q8WWM1 PRELIMINARY; PRT; 108 AA.
AC Q8WWM1;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE XAGE-5 protein.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RA Zandman A.J.W., van Kraats A.A., Weidle U.H., Ruiter D.J.,
RA Van Muijen G.N.P.;
RT "Expression profile and alignment of the XAGE family of cancer/testis
RT associated genes.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ318894; CAC83015.1; -.
SQ SEQUENCE 108 AA; 12077 MW; E3CD91E5C9241628 CRC64;

Query Match 30.8%; Score 199; DB 4; Length 108;
Best Local Similarity 42.4%; Pred. No. 4.1e-10;
Matches 50; Conservative 14; Mismatches 40; Indels 14; Gaps 4;

QY 1 MSNRGRSTYPRPRRYVEPPMIGMRPEQFSDEVEPA--TPEGEPATORQDPAAQAQEG 58
Db 1 MSNRGR---RYRPRCLRLAQLGPM-----LEPSVPEQQEPPPTESQDHTPGQR 49
QY 59 E-DEGASAGQPKPEAHSQEQGHPTQGCEDGPDGQEMDPNPEVKTTPERGEKQSQ 115
Db 50 EDQGAIEIQVPLEADLQELCQTKTGCGEDGSDPDVQKILPKSEQFKMPEGEGKQSQ 107

RESULT 7
Q8WWM0 PRELIMINARY; PRT; 69 AA.
AC Q8WWM0;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
```

```
DE XAGE-4 protein (Fragment).
GN XAGE-4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RA Zandman A.J.W., van Kraats A.A., Weidle U.H., Ruiter D.J.,
RA Van Muijen G.N.P.;
RT "Expression profile and alignment of the XAGE family of cancer/testis
RT associated genes.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ318895; CAC8302.1; -.
FT NON_TER 1
FT NON_TER 69
SQ SEQUENCE 69 AA; 7621 MW; 698F06B9E400E650 CRC64;

Query Match 19.3%; Score 124.5; DB 4; Length 69;
Best Local Similarity 33.7%; Pred. No. 0.00063;
Matches 30; Conservative 10; Mismatches 26; Indels 23; Gaps 3;

QY 19 PPENIGMRPEQFSDEVEPA--TPEGEPATORQDPAAQAQEGDEGASAGQPKPEAHSQ 76
Db 2 PPELIGM-----LEPSDEEPQOEPPPTESRDTPV-----PDLETDLQ 40
QY 77 EQGHPQTGCECDGPDGQEMDPNPEEVK 105
Db 41 ELSQKIGDECRDGPDDKGIQPNQSNLK 69

RESULT 8
O36421 PRELIMINARY; PRT; 1300 AA.
AC O36421;
DT 01-JAN-1998 (T-EMBLrel. 05, Created)
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Putative immediate early protein.
OS Alcelaphine herpesvirus 1 (wildbeest herpesvirus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=35252;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=CS00;
RX MEDLINE=37404659; PubMed=9261371;
RA Essner A., Pfanz R., Fleckenstein B.;
RT "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL J. Virol. 71:6517-6525(1997).
DR EMBL; AF005370; AAC58118.1; -.
DR InterPro; IPR000087; Colligen.
SQ SEQUENCE 1300 AA; 128183 MW; 40F9EFD244F34577 CRC64;

Query Match 18.4%; Score 119; DB 12; Length 1300;
Best Local Similarity 35.2%; Pred. No. 0.037;
Matches 45; Conservative 7; Mismatches 44; Indels 32; Gaps 7;

QY 17 VEPPEMIGMRPEQFSDEVEPA--TPEGEPATORQDPAAQAQEG-----E 59
Db 527 LEGPEGEGPEGPEGE--GPEGPEGPEGPEGPEGPEGPEGPEGPEGPEGPEGPE 585
QY 60 DEGASAGQPKPEAHSQEQGHPTQGCEDG-----PDQEMDPN-----PEVKT 106
Db 586 DEGEPEGPEGPEGPEGPEGPEGPEGPEGPEGPEGPEGPEGPEGPEGPEGPE 645
QY 107 PERGEKOS 114
Db 646 P-EGECOS 652

RESULT 9
O18563
```

QY	58	GDEGASAGQ-GPKPAHSQEQGHQTGCCEGDP--DGOEMDPNPNEVKTPE	108
Dd	478	PXPKPESPKQSPKQAPKEQPKPK-----PSPKQESSKQEPKPPEESPKE	526
 RESULT 11			
O9LJ64			
ID	O9LJ64	PRELIMINARY;	PRT; 956 AA.
AC	O9LJ64:		
DT	01-OCT-2000	(TREMBLrel. 15, Created)	
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)	
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)	
DE	Extensin protein-like.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=COLUMBIA;		
RA	Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=COLUMBIA;		
EX	MEDLINE=20363099; PubMed=10907853;		
RA	Nakamura Y.;		
RT	"Structural analysis of Arabidopsis thaliana chromosome 3. II.		
RT	Structure features of the regions of 4,251,695 bp covered by ninety P1		
RT	TAC and BAC clones.";		
RT	DNA Res. 7:217-221(2000).		
DR	EMBL; AP000735; BAB01698.1; .		
DR	InterPro; IPR001226; Flavodoxin.		
DR	InterPro; IPR001611; LRR.		
DR	InterPro; IPR003592; LRR out.		
DR	InterPro; IPR003882; Pistil extensin.		
DR	InterPro; IPR002965; P-rich_extensn.		
DR	Ffam; FFO560; LRR; 4.		
DR	PRINTS; PRO1217; PRICHEXTENSIN.		
DR	PRINTS; PRO1218; PSTLEXTENSIN.		
DR	SMART; SM00370; LRR; 5.		
DR	PROSITE; PS00201; FLAVODOXIN; UNKNOWN 1.		
SQ	SEQUENCE 956 AA; 102822 MW; E3AE135E63BB47C3 CRC64;		
 Query Match 18.0%; Score 116.5; DB 10; Length 956;			
Best Local Similarity 27.2%; Pred. No. 0.045;			
Matches 31; Conservative 22; Mismatches 40; Indels 21; Gaps			
QY	11	PPRPRVVEPPEMIGPM-----RPEQFSDEV-----EPATPSEGEPTAQODPAAQAQE	57
Dd	418	PFKKKEINPNLEEPSKXPESKQPQQSPKDETPTSHSNPKPKPESPKQTEQ	477
QY	58	GDEGASAGQ-GPKPAHSQEQGHQTGCCEGDP--DGOEMDPNPNEVKTPE	108
Dd	478	PXPKPESPKQSPKQAPKEQPKPK-----PSPKQESSKQEPKPPEESPKE	526
 RESULT 12			
O18564			
ID	O18564	PRELIMINARY;	PRT; 144 AA.
AC	O18564:		
DT	01-JAN-1998	(TREMBLrel. 05, Created)	
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	IGG and IgE immunoreactive antigen recognized by sera from patients		
DE	with strongyloidiasis (fragment).		
OS	Strongyloides stercoralis.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;		
OC	Panagrolaimoidea; Strongyloidea; Strongylidae;		
OX	NCBI_TaxID=6248;		
RN	[1]		

Search completed: April 7, 2003, 18:15:45
Job time : 36 secs

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2003, 18:13:10 ; Search time 35 Seconds
(without alignments)
441.631 Million cell updates/sec

Title: US-09-782-745-27
Perfect score: 646
Sequence: 1 MSWGRSTYRPRRYVPP.....DDPNBEVKTPEGEKQSQC 116

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :			
A. Geneseq_101002.*			
1:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*		
23:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	646	100.0	116	21	AAU83159
2	638	98.8	116	19	AAU47599
3	623.5	96.5	117	21	AAU83162
4	619.5	95.9	117	21	AAU83163
5	616.5	95.4	117	19	AAU47601
6	615.5	95.3	117	21	AAU83164
7	613	94.9	118	21	AAU83160
8	612.5	94.8	117	19	AAU47602
9	608.5	94.2	117	19	AAU47603
10	598	92.6	138	19	AAU47598

11	598	92.6	138	23	AAU84812
12	596.5	92.3	118	21	AAU83161
13	589.5	91.3	118	19	AAU47600
14	400	61.9	76	21	AAU802123
15	271.5	42.0	111	22	AAU39588
16	268.5	41.6	112	22	ABG05297
17	266	41.2	146	21	AAU80749
18	258	39.9	3541	23	AAU85130
19	249.5	38.6	106	22	AAU41374
20	229.5	35.5	111	22	AAU78785
21	229.5	35.5	111	22	AAU860500
22	229.5	35.5	111	22	ABU305753
23	224	34.7	117	22	AAU79769
24	173	26.8	30	23	AAU85017
25	163	25.2	30	23	AAU85013
26	162	25.1	30	23	AAU85016
27	160	24.8	30	23	AAU85012
28	156	24.1	30	23	AAU85014
29	156	24.1	30	23	AAU85015
30	155.5	24.1	60	22	ABG05299
31	141	21.8	115	22	AAU808583
32	132	20.4	249	22	ABG27048
33	120	18.6	30	23	AAU85018
34	120	18.6	102	21	AAU83158
35	120	18.6	102	21	AAU52998
36	120	18.6	102	23	AAU14575
37	113.5	17.6	87	21	AAU83167
38	109.5	17.0	111	22	AAU808581
39	108.5	16.8	79	21	AAU83169
40	104.5	16.2	507	14	AAU37682
41	104	16.1	89	20	AAU12447
42	103.5	16.0	281	21	AAU90256
43	102	15.8	2047	23	AAU75883
44	101	15.6	905	18	AAU31186
45	101	15.6	1135	18	AAU31185

ALIGNMENTS

RESULT 1
AAU83159
ID AAU83159 standard; Protein; 116 AA.

AC AAU83159;
DT 24-JUL-2000 (first entry)
DE GAGE2 polypeptide.

XX PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
cytotoxic T lymphocyte; immune response; antibody; drug delivery;
immunoconjugate.

OS Homo sapiens.
XX WO200012706-A1.
PN 09-MAR-2000.
XX 31-AUG-1999; 99WO-US20046.
XX 01-SEP-1998; 98US-0098993.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX Pastan I, Brinkmann U, Vasmatazis G, Lee B;
XX WPI; 2000-237869/20.

XX Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
lymphocyte response and for raising antibodies which can be used to

PT detect the presence of PAGE-4 in cell samples or body tissues
XX
XX
PS Disclosure; Figure 1a; 63pp; English.
XX
XX PAGE-4 is a gene preferentially expressed in normal male and female
CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC and placenta, as well as in prostate cancer, testicular cancer and
CC uterine cancer. This expression pattern makes it a target for
CC diagnosis and for vaccine based therapy of such neoplasms.
CC An isolated PAGE-4 peptide which induces a cytotoxic T
CC lymphocyte response when bound to a major histocompatibility complex
CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC against cells expressing PAGE-4 including cancer cells of the
CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC PAGE-4 peptide fragments can also be used in these compositions.
CC Antibodies against PAGE-4 and its peptide fragments can be used in
CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC samples or body tissues. The presence of PAGE-4 in tissues which are
CC not related to reproduction can be indicative of the spread of
CC cancerous reproductive tissue. PAGE-4 can also be used to raise
CC antibodies which are then used as the targeting group of
CC immunocjugates comprising toxins used in therapeutic applications.
CC This has applications for drug delivery systems. The PAGE
CC polypeptide shares sequence similarity with the GAGE and MAGE family
CC of proteins.
XX
SQ Sequence 116 AA;

Query Match 100.0%; Score 646; DB 21; Length 116;
Best Local Similarity 100.0%; Pred. No. 1e-52;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSWRGRSTYPRPRRYVEPPMIGMRPEQFSDEVEPATPEEGEPATQRODPAQAQSGED 60
Db 1 MSWRGRSTYPRPRRYVEPPMIGMRPEQFSDEVEPATPEEGEPATQRODPAQAQSGED 60
QY 61 EGASAGQPKPEAHSQEGHPQTGCECEDGPDQEMDPNPNPEVKTPPEGEKQSQ 116
Db 61 EGASAGQPKPEAHSQEGHPQTGCECEDGPDQEMDPNPNPEVKTPPEGEKQSQ 116

RESULT 2
AAW47599
ID AAW47599 standard; Protein; 116 AA.
XX
AC AAW47599;
XX
XX 30-JUL-1998 (first entry)
XX
DE GAGE-2 tumour rejection antigen precursor.
XX
XX GAGE tumour rejection antigen precursor; TRAP; tumour;
KW diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW HLA-typing assay.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 38 /note= "Ala encoded by GAG"
FT Misc-difference 39 /note= "Thr encoded by CCT"
FT
XX
PN WO9749417-Al.
XX
XX 31-DEC-1997.
XX
XX 23-JUN-1997; 97WO-US10850.
XX
XX 24-JUN-1996; 96US-0669161.
PR
XX (LUDW-) LUDWIG INST CANCER RES.

XX
PI Boon-Talleur T, Debacker O, Van Den Eynde B;
XX
XX WPI; 1998-076905/07.
DR N-PSDB; AAV18717.
XX
XX Isolated nucleic acid encoding GAGE tumour rejection antigen
PT precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT diagnose melanomas
XX
XX Example 13; Fig 5; 60pp; English.
XX
XX The present sequence represents a GAGE-2 tumour rejection antigen
CC precursor (TRAP). The protein is expressed in a number of tumours. In
CC contrast the only normal tissue which expresses GAGE TRAP protein
CC is testis. Several GAGE TRAPs have been identified (see AAV18717-21).
CC The major difference between these proteins and GAGE-1 is the absence
CC of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
CC TRAP sequence. The rest of the sequences show mismatches at various
CC position, with the exception of GAGE-3 whose 5' end is totally
CC different from the other GAGE cDNAs for the first 112 bases. This
CC region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC The antigens can be used to diagnose melanomas, characterised by
CC expression of a TRAP or presentation of a tumour rejection antigen.
CC Antigens shed into blood or urine can be observed and then used to
CC confirm a diagnosis of melanoma using cytolytic T cell clone
CC proliferation methodologies. Other uses for the processed peptides,
CC include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ Sequence 116 AA;
Query Match 98.8%; Score 638; DB 19; Length 116;
Best Local Similarity 99.1%; Pred. No. 5.5e-52;
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSWRGRSTYPRPRRYVEPPMIGMRPEQFSDEVEPATPEEGEPATQRODPAQAQSGED 60
Db 1 MSWRGRSTYPRPRRYVEPPMIGMRPEQFSDEVEPATPEEGEPATQRODPAQAQSGED 60
QY 61 EGASAGQPKPEAHSQEGHPQTGCECEDGPDQEMDPNPNPEVKTPPEGEKQSQ 116
Db 61 EGASAGQPKPEAHSQEGHPQTGCECEDGPDQEMDPNPNPEVKTPPEGEKQSQ 116
RESULT 3
AAV83162
ID AAV83162 standard; Protein; 117 AA.
XX
AC AAV83162;
XX
XX 24-JUL-2000 (first entry)
XX
DE GAGE4 polypeptide.
XX
XX PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW cytotoxic T lymphocyte; immune response; antibody; drug delivery;
KW immunoconjugate.
XX
XX Homo sapiens.
XX
XX WO200012706-Al.
FN
XX
XX 09-MAR-2000.
XX
XX 31-AUG-1999; 99WO-US20046.
XX
XX 01-SEP-1998; 98US-0098993.
PR
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Pastan I, Brinkmann U, Vasmatzis G, Lee B;
PI
XX

DR WPI; 2000-237869/20.
 XX Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
 PT lymphocyte response and for raising antibodies which can be used to
 PT detect the presence of PAGE-4 in cell samples or body tissues
 XX
 PS Disclosure; Figure 1a; 63pp; English.
 XX
 CC PAGE-4 is a gene preferentially expressed in normal male and female
 CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus
 CC and placenta, as well as in prostate cancer, testicular cancer and
 CC uterine cancer. This expression pattern makes it a target for
 CC diagnosis and for vaccine based therapy of such neoplasms.
 CC An isolated PAGE-4 peptide which induces a cytotoxic T
 CC lymphocyte response when bound to a major histocompatibility complex
 CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in
 CC immunogenic compositions to raise a cytotoxic T lymphocyte response
 CC against cells expressing PAGE-4 including cancer cells of the
 CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
 CC PAGE-4 peptide fragments can also be used in these compositions.
 CC Antibodies against PAGE-4 and its peptide fragments can be used in
 CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
 CC samples or body tissues. The presence of PAGE-4 in tissues which are
 CC not related to reproduction can be indicative of the spread of
 CC cancerous reproductive tissue. PAGE-4 can also be used to raise
 CC antibodies which are then used as the targeting group of
 CC immunoconjugates comprising toxins used in therapeutic applications.
 CC This has applications for drug delivery systems. The PAGE
 CC polypeptide shares sequence similarity with the GAGE and MAGE family
 CC of proteins.
 XX
 SQ Sequence 117 AA;

Query Match 96.5%; Score 623.5; DB 21; Length 117;
 Best Local Similarity 97.4%; Pred. No. 1.2e-50;
 Matches 114; Conservative 1; Mismatches 1; Indels 1; Gaps 1;
 Qy 1 MSWRGRST-YRPRRYVEPPMIGMPRPEQFSDVEVPATPEGEPAQRDPAAAEQGE 59
 Db 1 MSWRGRSTYRPRRYVQPPMIGMPRPEQFSDVEVPATPEGEPAQRDPAAAEQGE 60
 Qy 60 DEGASAGQPKPEAHSEQGHPTGCECEDGPDGQEMDPPNPPEVKTPEGEKQSQ 116
 Db 61 DEGASAGQPKPEADSQEQGHPTGCECEDGPDGQEMDPPNPPEVKTPEGEKQSQ 117

RESULT 4
 AAY83163
 ID AAY83163 standard; Protein; 117 AA.
 XX
 AC AAY83163;
 XX
 DT 24-JUL-2000 (first entry)
 XX
 DE GAGE5 polypeptide.
 XX
 KW PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
 KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
 KW cytotoxic T lymphocyte; immune response; antibody; drug delivery;
 KW immunoconjugate.
 XX
 OS Homo sapiens.
 XX
 PN WO200012706-A1.
 XX
 PD 09-MAR-2000.
 XX
 PF 31-AUG-1999; 99WO-US20046.
 XX
 PR 01-SEP-1998; 98US-0098993.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Pastan I, Brinkmann U, Vasmatazis G, Lee B;
 XX WPI; 2000-237869/20.
 XX Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
 PT lymphocyte response and for raising antibodies which can be used to
 PT detect the presence of PAGE-4 in cell samples or body tissues
 XX
 PS Disclosure; Figure 1a; 63pp; English.
 XX
 CC PAGE-4 is a gene preferentially expressed in normal male and female
 CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus
 CC and placenta, as well as in prostate cancer, testicular cancer and
 CC uterine cancer. This expression pattern makes it a target for
 CC diagnosis and for vaccine based therapy of such neoplasms.
 CC An isolated PAGE-4 peptide which induces a cytotoxic T
 CC lymphocyte response when bound to a major histocompatibility complex
 CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in
 CC immunogenic compositions to raise a cytotoxic T lymphocyte response
 CC against cells expressing PAGE-4 including cancer cells of the
 CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
 CC PAGE-4 peptide fragments can also be used in these compositions.
 CC Antibodies against PAGE-4 and its peptide fragments can be used in
 CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
 CC samples or body tissues. The presence of PAGE-4 in tissues which are
 CC not related to reproduction can be indicative of the spread of
 CC cancerous reproductive tissue. PAGE-4 can also be used to raise
 CC antibodies which are then used as the targeting group of
 CC immunoconjugates comprising toxins used in therapeutic applications.
 CC This has applications for drug delivery systems. The PAGE
 CC polypeptide shares sequence similarity with the GAGE and MAGE family
 CC of proteins.
 XX
 SQ Sequence 117 AA;

Query Match 95.9%; Score 619.5; DB 21; Length 117;
 Best Local Similarity 96.6%; Pred. No. 2.9e-50;
 Matches 113; Conservative 2; Mismatches 1; Indels 1; Gaps 1;
 Qy 1 MSWRGRST-YRPRRYVEPPMIGMPRPEQFSDVEVPATPEGEPAQRDPAAAEQGE 59
 Db 1 MSWRGRSTYRPRRYVQPPMIGMPRPEQFSDVEVPATPEGEPAQRDPAAAEQGE 60
 Qy 60 DEGASAGQPKPEAHSEQGHPTGCECEDGPDGQEMDPPNPPEVKTPEGEKQSQ 116
 Db 61 DEGASAGQPKPEADSQEQGHPTGCECEDGPDGQEMDPPNPPEVKTPEGEKQSQ 117

RESULT 5
 AAW47601
 ID AAW47601 standard; Protein; 117 AA.
 XX
 AC AAW47601;
 XX
 DT 30-JUL-1998 (first entry)
 XX
 DE GAGE-4 tumour rejection antigen precursor.
 XX
 KW GAGE tumour rejection antigen precursor; TRAP; tumour;
 KW diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
 KW HLA-typing assay.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 39 /note= "Ala encoded by GAG"
 FT Misc-difference 40 /note= "Thr encoded by CCT"
 XX
 PN WO9749417-A1.
 XX
 PD 31-DEC-1997.

XX 23-JUN-1997; 97WO-US10850.
 XX 24-JUN-1996; 96US-0669161.
 XX (LUDW-) LUDWIG INST CANCER RES.
 XX Boon-Falleur T, Debacker O, Van Den Eynde B;
 PI WPI; 1998-076905/07.
 XX N-PSDB; AAV18719.
 XX Isolated nucleic acid encoding GAGE tumour rejection antigen
 PT precursor - processed by HLA-Cw6 molecules into peptides, useful to
 PT diagnose melanomas
 XX Example 13; Fig 5; 60pp; English.
 PS
 XX The present sequence represents a GAGE-4 tumour rejection antigen
 CC precursor (TRAP). The protein is expressed in a number of tumours. In
 CC contrast the only normal tissue which expresses GAGE TRAP protein
 CC is testis. Several GAGE TRAPS have been identified (see AAV18717-21).
 CC The major difference between these proteins and GAGE-1 is the absence
 CC of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
 CC TRAP sequence. The rest of the sequences show mismatches at various
 CC position, with the exception of GAGE-3 whose 5' end is totally
 CC different from the other GAGE cDNAs for the first 112 bases. This
 CC region of GAGE-3 CDNA contains a long repeat and a hairpin structure.
 CC The antigens can be used to diagnose melanomas, characterised by
 CC expression of a TRAP or presentation of a tumour rejection antigen.
 CC Antigens shed into blood or urine can be observed and then used to
 CC confirm a diagnosis of melanoma using cytolytic T cell clone
 CC proliferation methodologies. Other uses for the processed peptides,
 CC include HLA-typing assays for, e.g. skin graft or organ transplants.
 XX
 XX Sequence 117 AA;
 SQ
 Query Match 95.4%; Score 616.5; DB 19; Length 117;
 Best Local Similarity 96.6%; Pred. No. 5.6e-50;
 Matches 113; Conservative 1; Mismatches 2; Indels 1; Gaps 1;
 QY 1 MSWRGRSTYR-PPRRYVPEPPMIGMRPQFSDVEVPATPEGEPTATQRODPAQAQGE 59
 DB 1 MSWRGRSTYRPPRRYVPEPPMIGMRPQFSDVEVPATPEGEPTATQRODPAQAQGE 60
 QY 60 DEGASAGQPKPEAHSQEQGHPTGCECDGPDQEMDPPNPPEVKTPPEGEKQSQ 116
 DB 61 DEGASAGQPKPEADSQEQGHPTGCECDGPDQEMDPPNPPEVKTPPEGEKQSQ 117
 RESULT 6
 AAY83164
 ID AAY83164 standard; Protein; 117 AA.
 XX AC AAY83164;
 XX 24-JUL-2000 (first entry)
 XX GAGE6 polypeptide.
 XX
 XX PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
 KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
 KW cytotoxic T lymphocyte; immune response; antibody; drug delivery;
 KW immunoconjugate.
 XX
 XX Homo sapiens.
 XX
 XX WO200012706-A1.
 XX
 XX 09-MAR-2000.
 XX
 XX 1-AUG-1999; 99WO-US20046.
 XX

PR 01-SEP-1998; 98US-0098993.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Pastan I, Brinkmann U, Vasmatazis G, Lee B;
 PI WPI; 2000-237869/20.
 XX
 XX Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
 PT lymphocyte response and for raising antibodies which can be used to
 PT detect the presence of PAGE-4 in cell samples or body tissues
 XX
 XX Disclosure; Figure 1a; 63pp; English.
 XX
 CC PAGE-4 is a gene preferentially expressed in normal male and female
 CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus
 CC and placenta, as well as in prostate cancer, testicular cancer and
 CC uterine cancer. This expression pattern makes it a target for
 CC diagnosis and for vaccine based therapy of such neoplasms.
 CC An isolated PAGE-4 peptide which induces a cytotoxic T
 CC lymphocyte response when bound to a major histocompatibility complex
 CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in
 CC immunogenic compositions to raise a cytotoxic T lymphocyte response
 CC against cells expressing PAGE-4 including cancer cells of the
 CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
 CC PAGE-4 peptide fragments can also be used in these compositions.
 CC Antibodies against PAGE-4 and its peptide fragments can be used in
 CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
 CC samples or body tissues. The presence of PAGE-4 in tissues which are
 CC not related to reproduction can be indicative of the spread of
 CC cancerous reproductive tissue. PAGE-4 can also be used to raise
 CC antibodies which are then used as the targeting group of
 CC immunoconjugates comprising toxins used in therapeutic applications.
 CC This has applications for drug delivery systems. The PAGE
 CC polypeptide shares sequence similarity with the GAGE and MAGE family
 CC of proteins.
 XX
 XX Sequence 117 AA;
 SQ
 Query Match 95.3%; Score 615.5; DB 21; Length 117;
 Best Local Similarity 95.7%; Pred. No. 6.9e-50;
 Matches 112; Conservative 3; Mismatches 1; Indels 1; Gaps 1;
 QY 1 MSWRGRSTYRPPRRYVPEPPMIGMRPQFSDVEVPATPEGEPTATQRODPAQAQGE 59
 DB 1 MSWRGRSTYRPPRRYVPEPPMIGMRPQFSDVEVPATPEGEPTATQRODPAQAQGE 60
 QY 60 DEGASAGQPKPEAHSQEQGHPTGCECDGPDQEMDPPNPPEVKTPPEGEKQSQ 116
 DB 61 DEGASAGQPKPEADSQEQGHPTGCECDGPDQEMDPPNPPEVKTPPEGEKQSQ 117
 RESULT 7
 AAY83160
 ID AAY83160 standard; Protein; 118 AA.
 XX AC AAY83160;
 XX 24-JUL-2000 (first entry)
 XX GAGE1 polypeptide.
 XX
 XX PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
 KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
 KW cytotoxic T lymphocyte; immune response; antibody; drug delivery;
 KW immunoconjugate.
 XX
 XX Homo sapiens.
 XX
 XX WO200012706-A1.
 XX
 XX 09-MAR-2000.
 XX

PF 31-AUG-1999; 99WO-US20046.
 XX 01-SEP-1998; 98US-0098993.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Pastan I, Brinkmann U, Vasmatazis G, Lee B;
 XX WPI; 2000-237869/20.
 XX Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
 PT lymphocyte response and for raising antibodies which can be used to
 PT detect the presence of PAGE-4 in cell samples or body tissues
 XX Disclosure; Figure 1a; 63pp; English.
 XX PAGE-4 is a gene preferentially expressed in normal male and female
 CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus
 CC and placenta, as well as in prostate cancer, testicular cancer and
 CC uterine cancer. This expression pattern makes it a target for
 CC diagnosis and for vaccine based therapy of such neoplasms.
 CC An isolated PAGE-4 peptide which induces a cytotoxic T
 CC lymphocyte response when bound to a major histocompatibility complex
 CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in
 CC against cells expressing PAGE-4 including cancer cells of the
 CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
 CC PAGE-4 peptide fragments can also be used in these compositions.
 CC Antibodies against PAGE-4 and its peptide fragments can be used in
 CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
 CC samples or body tissues. The presence of PAGE-4 in tissues which are
 CC not related to reproduction can be indicative of the spread of
 CC cancerous reproductive tissue. PAGE-4 can also be used to raise
 CC antibodies which are then used as the targeting group of
 CC immunoconjugates comprising toxins used in therapeutic applications.
 CC This has applications for drug delivery systems. The PAGE
 CC polypeptide shares sequence similarity with the GAGE and MAGE family
 CC of proteins.
 XX SQ Sequence 118 AA;
 Query Match 94.9%; Score 613; DB 21; Length 118;
 Best Local Similarity 97.4%; Pred. No. 1.2e-49;
 Matches 111; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSWRGRSTYRPRRYVPEPMIGMRPEQFSDEVEPATPEGEPTAQDPAQAQEGED 60
 DB 1 MSWRGRSTYRPRRYVPEPMIGMRPEQFSDEVEPATPEGEPTAQDPAQAQEGED 60
 QY 61 EGASAGQGPKPEAHSQEQGHPTQCECEDGPDGQMDPNPEVKTPEGEKQS 114
 DB 61 EGASAGQGPKPEAHSQEQGHPTQCECEDGPDGQMDPNPEVKTPEGEKQS 114
 RESULT 8
 AAW47602
 ID AAW47602 standard; Protein; 117 AA.
 XX AAW47602;
 XX 30-JUL-1998 (first entry)
 XX GAGE-5 tumour rejection antigen precursor.
 XX GAGE tumour rejection antigen precursor; TRAP; tumour;
 KW diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
 KW HLA-typing assay.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FT Misc-difference 39 /note= "Ala encoded by GAG"

FT Misc-difference 40 /note= "Thr encoded by CCT"
 XX WO9749417-A1.
 XX 31-DEC-1997.
 XX 23-JUN-1997; 97WO-US10850.
 XX 24-JUN-1996; 96US-0669161.
 XX (LUDW-) LUDWIG INST CANCER RES.
 XX Boon-Falleur T, Debacker O, Van Den Eynde B;
 XX WPI; 1998-076905/07.
 XX N-PSDB; AAV18720.
 XX Isolated nucleic acid encoding GAGE tumour rejection antigen
 PT precursor - processed by HLA-Cw6 molecules into peptides, useful to
 PT diagnose melanomas
 XX Example 13; Fig 5; 60pp; English.
 XX The present sequence represents a GAGE-5 tumour rejection antigen
 CC precursor (TRAP). The protein is expressed in a number of tumours. In
 CC contrast the only normal tissue which expresses GAGE TRAP protein
 CC is testis. Several GAGE TRAPS have been identified (see AAV18717-21).
 CC The major difference between these proteins and GAGE-1 is the absence
 CC of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
 CC TRAP sequence. The rest of the sequences show mismatches at various
 CC positions, with the exception of GAGE-3 whose 5' end is totally
 CC different from the other GAGE cDNAs for the first 112 bases. This
 CC region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
 CC The antigens can be used to diagnose melanomas, characterised by
 CC expression of a TRAP or presentation of a tumour rejection antigen.
 CC Antigens shed into blood or urine can be observed and then used to
 CC confirm a diagnosis of melanoma using cytolytic T cell clone
 CC proliferation methodologies. Other uses for the processed peptides,
 CC include HLA-typing assays for, e.g. skin graft or organ transplants.
 XX SQ Sequence 117 AA;
 Query Match 94.8%; Score 612.5; DB 19; Length 117;
 Best Local Similarity 95.7%; Pred. No. 1.3e-49;
 Matches 112; Conservative 2; Mismatches 2; Indels 1; Gaps 1;
 QY 1 MSWRGRSTYR-PRRYVPEPMIGMRPEQFSDEVEPATPEGEPTAQDPAQAQEGE 59
 DB 1 MSWRGRSTYRPRRYVQPEVIGMRPEQFSDEVEPATPEGEPTAQDPAQAQEGE 60
 QY 60 DEGASAGQGPKPEAHSQEQGHPTQCECEDGPDGQMDPNPEVKTPEGEKQS 116
 DB 61 DEGASAGQGPKPEAHSQEQGHPTQCECEDGPDGQMDPNPEVKTPEGEKQS 117
 RESULT 9
 AAW47603
 ID AAW47603 standard; Protein; 117 AA.
 XX AAW47603;
 XX 30-JUL-1998 (first entry)
 XX GAGE-6 tumour rejection antigen precursor.
 XX GAGE tumour rejection antigen precursor; TRAP; tumour;
 KW diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
 KW HLA-typing assay.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FT Misc-difference 39 /note= "Ala encoded by GAG"

```

FT Misc-difference 39 /note= "Ala encoded by GAG"
FT Misc-difference 40 /note= "Thr encoded by CCT"
XX WO9749417-A1.
XX 31-DEC-1997.
XX 23-JUN-1997; 97WO-US10850.
XX 24-JUN-1996; 96US-0669161.
XX (LUDW-) LUDWIG INST CANCER RES.
XX Boon-Falleur T, Debacker O, Van Den Eynde B;
XX WPI; 1998-076905/07.
XX N-PSDB; AAV18721.
XX Isolated nucleic acid encoding GAGE tumour rejection antigen
XX precursor - processed by HLA-Cw6 molecules into peptides, useful to
XX diagnose melanomas
XX Example 13; Fig 5; 60pp; English.
XX The present sequence represents a GAGE-6 tumour rejection antigen
XX precursor (TRAP). The protein is expressed in a number of tumours. In
XX contrast the only normal tissue which expresses GAGE TRAP protein
XX is testis. Several GAGE TRAPS have been identified (see AAV18717-21).
XX The major difference between these proteins and GAGE-1 is the absence
XX of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
XX TRAP sequence. The rest of the sequences show mismatches at various
XX positions, with the exception of GAGE-3 whose 5' end is totally
XX different from the other GAGE cDNAs for the first 112 bases. This
XX region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
XX The antigens can be used to diagnose melanomas, characterised by
XX expression of a TRAP or presentation of a tumour rejection antigen.
XX Antigens shed into blood or urine can be observed and then used to
XX confirm a diagnosis of melanoma using cytolytic T cell clone
XX proliferation methodologies. Other uses for the processed peptides,
XX include HLA-typing assays for, e.g. skin graft or organ transplants.
XX Sequence 117 AA;

Query Match 94.2%; Score 608.5; DB 19; Length 117;
Best Local Similarity 94.9%; Pred. No. 3.1e-49;
Matches 111; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 1 MSWRGRSTYR-PRRRYVEPPMIGMPRPEQFSDEVEPATPEEGEPATQRQDPAAQEGE 59
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
60 1 MSWRGRSTYR-PRRRYVEPPMIGMPRPEQFSDEVEPATPEEGEPATQRQDPAAQEGE 60
QY 60 DEGASAGQGPKEAHSQEQGHFQTGCECEDGPDGQEMDPPNPERVKTPEEGEKQSQC 116
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 DEGASAGQGPKEADSDQEQGHFQTGCECEDGPDGQEMDPPNPERVKTPEEGEKQSQC 117

RESULT 10
AAW47598
ID AAW47598 standard; Protein; 138 AA.
XX AC AAW47598;
XX 30-JUL-1998 (first entry)
XX GAGE-1 tumour rejection antigen precursor.
XX GAGE tumour rejection antigen precursor; TRAP; tumour;
XX diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
XX HLA-typing assay.
XX Homo sapiens.
OS

```

```

XX Key Location/Qualifiers
FH Misc-difference 39 /note= "encoded by GAG"
FT Misc-difference 40 /note= "encoded by CCT"
XX WO9749417-A1.
XX 31-DEC-1997.
XX 23-JUN-1997; 97WO-US10850.
XX 24-JUN-1996; 96US-0669161.
XX (LUDW-) LUDWIG INST CANCER RES.
XX Boon-Falleur T, Debacker O, Van Den Eynde B;
XX WPI; 1998-076905/07.
XX N-PSDB; AAV05540.
XX Isolated nucleic acid encoding GAGE tumour rejection antigen
XX precursor - processed by HLA-Cw6 molecules into peptides, useful to
XX diagnose melanomas
XX Example 7; Fig 5; 60pp; English.
XX The present sequence represents a GAGE-1 tumour rejection antigen
XX precursor (TRAP). The protein is expressed in a number of tumours. In
XX contrast the only normal tissue which expresses GAGE TRAP protein
XX is testis. Several GAGE TRAPS have been identified (see AAV18717-21).
XX The major difference between these proteins and GAGE-1 is the absence
XX of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
XX TRAP sequence. The rest of the sequences show mismatches at various
XX positions, with the exception of GAGE-3 whose 5' end is totally
XX different from the other GAGE cDNAs for the first 112 bases. This
XX region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
XX The antigens can be used to diagnose melanomas, characterised by
XX expression of a TRAP or presentation of a tumour rejection antigen.
XX Antigens shed into blood or urine can be observed and then used to
XX confirm a diagnosis of melanoma using cytolytic T cell clone
XX proliferation methodologies. Other uses for the processed peptides,
XX include HLA-typing assays for, e.g. skin graft or organ transplants.
XX Sequence 138 AA;

Query Match 92.6%; Score 598; DB 19; Length 138;
Best Local Similarity 99.1%; Pred. No. 3.5e-48;
Matches 108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSWRGRSTYR-PRRRYVEPPMIGMPRPEQFSDEVEPATPEEGEPATQRQDPAAQEGE 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 MSWRGRSTYR-PRRRYVEPPMIGMPRPEQFSDEVEPATPEEGEPATQRQDPAAQEGE 60
QY 61 EGASAGQGPKEAHSQEQGHFQTGCECEDGPDGQEMDPPNPERVKTPEE 109
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 EGASAGQGPKEADSDQEQGHFQTGCECEDGPDGQEMDPPNPERVKTPEE 109

RESULT 11
AAU84812
ID AAU84812 standard; Protein; 138 AA.
XX AC AAU84812;
XX 08-MAY-2002 (first entry)
XX Human GAGE-1 consensus sequence.
XX Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
XX viral infection; human immunodeficiency virus; melanoma;
XX bacterial infection; Salmonella; Legionella; parasitic infection;
XX Homo sapiens.
OS

```

KW Trypanosoma; Toxoplasma; Giardia.

XX Homo sapiens.

OS WO200190197-A1.

PN 29-NOV-2001.

PD 25-MAY-2001; 2001WO-AU00622.

PF 26-MAY-2000; 2000AU-0007761.

PR (AUSU) UNIV AUSTRALIAN NAT.

PA Thomson SA, Ramshaw IA;

PI WPI; 2002-147575/19.

DR New synthetic polypeptides having several different segments of at

XX least one parent polypeptide linked together differently compared to

PT the linkage in the parent polypeptide, for inducing immune response

PT against a pathogen or cancer

XX Example 3; Fig 27; 364pp; English.

PS The invention relates to a new synthetic polypeptide (I) comprising

XX several different segments of at least one parent polypeptide linked

CC together in a different relationship relative to their linkage in the

CC parent polypeptide to impede, abrogate or otherwise alter at least one

CC function associated with the parent polypeptide and for inducing an

CC immune response against a pathogen or cancer. Also included are a

CC synthetic polynucleotide encoding and a computer system for

CC designing the synthetic polypeptides. The synthetic polypeptides and

CC polynucleotides are referred to as a vaccine. The synthetic polypeptides

CC useful for modulating immune responses preferably directed against a

CC pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix,

CC colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone

CC liver, oesophagus, brain, testicle, uterus), as potentiating agents.

CC Compositions comprising the polypeptide may be used in the treatment or

CC prophylaxis against viral (such as infections caused by HIV (human

CC immunodeficiency virus), hepatitis, influenza, Japanese encephalitis

CC virus, Epstein-Barr virus and respiratory syncytial virus), bacterial

CC (e.g., infections caused by Neisseria, Meningococcus, Haemophilus,

CC Salmonella, Streptococcus, Legionella and Mycobacterium) or parasitic

CC (e.g., infections caused by Plasmodium, Schistosoma, Leishmania,

CC Trypanosoma, Toxoplasma and Giardia) infections. The present

CC sequence is a consensus sequence for a parent protein used to design a

XX vaccine of the invention.

SQ Sequence 138 AA;

Query Match 92.8%; Score 598; DB 23; Length 138;
Best Local Similarity 99.1%; Pred. No. 3.5e-48;
Matches 108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSWRGSTRYPRRRYVEPPMIGMPMRPEQFSDEVEPATPEEGEPATQRQDPAAQAQEGED 60

DB 1 MSWRGSTRYPRRRYVEPPMIGMPMRPEQFSDEVEPATPEEGEPATQRQDPAAQAQEGED 60

QY 61 EGASAGGPKPEAHSQEQGHPTQCECEDGPDGQEMDPPNPEVKTPPEE 109

DB 61 EGASAGGPKPEADSOQGHPTQCECEDGPDGQEMDPPNPEVKTPPEE 109

RESULT 12

AA483161

ID AAY83161 standard; Protein; 118 AA.

XX AAY83161;

XX 24-JUL-2000 (first entry)

XX GAGE3 polypeptide.

XX

KW

OS

PN

PD

PF

PR

PA

PI

DR

XX

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

PAGE-4; MAGE; reproduction; testis; prostate; fallopian tube; uterus; placenta; cancer; major histocompatibility complex; MHC; CTL; cytotoxic T lymphocyte; immune response; antibody; drug delivery; immunoconjugate.

XX Homo sapiens.

XX WO200012706-A1.

XX 09-MAR-2000.

XX 31-AUG-1999; 99WO-US20046.

XX 01-SEP-1998; 98US-0098993.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Paetan I, Brinkmann U, Vasmatazis G, Lee B;

XX WPI; 2000-237869/20.

XX Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T

XX lymphocyte response and for raising antibodies which can be used to

XX detect the presence of PAGE-4 in cell samples or body tissues

XX Disclosure; Figure 1a; 63pp; English.

XX PAGE-4 is a gene preferentially expressed in normal male and female

XX reproductive tissues e.g. prostate, testis, fallopian tube, uterus

XX and placenta, as well as in prostate cancer, testicular cancer and

XX uterine cancer. This expression pattern makes it a target for

XX diagnosis and for vaccine based therapy of such neoplasms.

XX An isolated PAGE-4 peptide which induces a cytotoxic T

XX lymphocyte response when bound to a major histocompatibility complex

XX (MHC) class I molecule or the isolated PAGE-4 protein can be used in

XX immunogenic compositions to raise a cytotoxic T lymphocyte response

XX against cells expressing PAGE-4 including cancer cells of the

XX prostate, uterus and testis. The nucleic acids encoding PAGE-4 or

XX PAGE-4 peptide fragments can also be used in these compositions.

XX Antibodies against PAGE-4 and its peptide fragments can be used in

XX detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell

XX samples or body tissues. The presence of PAGE-4 in tissues which are

XX not related to reproduction can be indicative of the spread of

XX cancerous reproductive tissue. PAGE-4 can also be used to raise

XX antibodies which are then used as the targeting group of

XX immunoconjugates comprising toxins used in therapeutic applications.

XX This has applications for drug delivery systems. The PAGE

XX polypeptide shares sequence similarity with the GAGE and MAGE family

XX of proteins.

XX SQ Sequence 118 AA;

Query Match 92.3%; Score 596.5; DB 21; Length 118;
Best Local Similarity 95.6%; Pred. No. 4.1e-48;
Matches 109; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

QY 4 RGRSTYPRRRYVEPPMIGMPMRPEQFSDEVEPATPEEGEPATQRQDPAAQAQEGED 62

DB 5 RKGSTYPRRRYVEPPMIGMPMRPEQFSDEVEPATPEEGEPATQRQDPAAQAQEGED 64

QY 63 ASAGGPKPEAHSQEQGHPTQCECEDGPDGQEMDPPNPEVKTPPEGEKQSQC 116

DB 65 ASAGGPKPEADSOQGHPTQCECEDGPDGQEMDPPNPEVKTPPEGEKQSQC 118

RESULT 13

AA47600

ID AAW47600 standard; Protein; 118 AA.

XX AAW47600;

XX 30-JUL-1998 (first entry)

XX DE GAGE-3 tumour rejection antigen precursor.
XX DE GAGE tumour rejection antigen precursor; TRAP; tumour;
XX KW diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
XX KW HLA-typing assay.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Misc-difference 40 /note= "ala encoded by GAG"
FT Misc-difference 41 /note= "Thr encoded by CCT"
FT
XX WO9749417-A1.
XX 31-DEC-1997.
XX 23-JUN-1997; 97WO-US10850.
XX 24-JUN-1996; 96US-0669161.
XX (LUDW-) LUDWIG INST CANCER RES.
XX Boon-Falleur T, Debacker O, Van Den Eynde B:
XX WPI; 1998-076905/07.
XX N-PSDB; AAV18718.
XX Isolated nucleic acid encoding GAGE tumour rejection antigen
XX precursor - processed by HLA-Cw6 molecules into peptides, useful to
XX diagnose melanomas
XX Example 13; Fig 5; 60pp; English.
XX The present sequence represents a GAGE-3 tumour tumour rejection antigen
XX precursor (TRAP). The protein is expressed in a number of tumours. In
XX contrast the only normal tissue which expresses GAGE TRAP protein
XX is testis. Several GAGE TRAPs have been identified (see AAV18717-21).
XX The major difference between these proteins and GAGE-1 is the absence
XX of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
XX TRAP sequence. The rest of the sequences show mismatches at various
XX positions, with the exception of GAGE-3 whose 5' end is totally
XX different from the other GAGE cDNAs for the first 112 bases. This
XX region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
XX The antigens can be used to diagnose melanomas, characterised by
XX expression of a TRAP or presentation of a tumour rejection antigen.
XX Antigens shed into blood or urine can be observed and then used to
XX confirm a diagnosis of melanoma using cytolytic T cell clone
XX proliferation methodologies. Other uses for the processed peptides,
XX include HLA-typing assays for, e.g. skin graft or organ transplants.
XX Sequence 118 AA;
Query Match 91.3%; Score 589.5; DB 19; Length 118;
Best Local Similarity 94.7%; Pred. No. 1.8e-47;
Matches 108; Conservative 3; Mismatches 2; Indels 1; Gaps 1;
Qy 4 RGRSTYR-PRPRRYVPPMIGMPMPQFSDEVEPATPEEGEPATQRQDPAAQAQGEDEG 62
Db 5 RGSSTVYVPRPRRYVQPEVIGMPMPQFSDEVEPATPEEGEPATQRQDPAAQAQGEDEG 64
Qy 63 ASAGGPKPEAHSHOQGHQPTGCECDGPGQEMDPNPPEVKTPPEGEKQSQ 116
Db 65 ASAGGPKPEADSEQGHQPTGCECDGPGQEMDPNPPEVKTPPEGEKQSQ 118
RESULT 14
AAG02123
ID AAG02123 standard; Protein; 76 AA.
XX
AC AAG02123;

XX 06-OCT-2000 (first entry)
XX Human secreted protein, SEQ ID NO: 6204.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PF 21-FEB-2000; 2000EP-0200610.
XX PR 26-FEB-1999; 99US-0122487.
XX (GEST) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX N-PSDB; AAC02129.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 13; SEQ ID 6204; 71pp + CD-ROM; English.
XX The present sequence is a polypeptide encoded by one of a large number
XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
XX were prepared from total human RNAs or polyA+ RNAs derived from 30
XX different tissues. EST sequences usually correspond mainly to the 3'
XX untranslated region (UTR) of the mRNA because they are often obtained
XX from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
XX isolating cDNA sequences derived from the 5' ends of mRNAs and even in
XX those cases where longer cDNA sequences have been obtained, the full 5'
XX UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
XX ends and can therefore be used to obtain full length cDNAs and genomic
XX DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
XX chromosome mapping procedures. They are used to obtain upstream
XX regulatory sequences and to design expression and secretion vectors.
XX Sequence 76 AA;
Query Match 61.9%; Score 400; DB 21; Length 76;
Best Local Similarity 97.3%; Pred. No. 4.9e-30;
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 22 MIGMPMPQFSDEVEPATPEEGEPATQRQDPAAQAQGEDEGASAGGPKPEAHSHOQGH 81
Db 1 MIGMPMPQFSDEVEPATPEEGEPATQRQDPAAQAQGEDEGASAGGPKPEAHSHOQGH 60
Qy 82 QTGCECEDGPGQEM 96
Db 61 QTGCECEDGPGQEM 75
RESULT 15
AAM39588
ID AAM39588 standard; Protein; 111 AA.
XX AAM39588;
XX 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 2733.
XX Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX KW peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0498725.

XX 25-APR-2000; 2000US-052317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Auandi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI58744.

XX Novel nucleic acids and polypeptides, useful for treating disorders

XX such as central nervous system injuries -

XX Example 4; SEQ ID NO 2733; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 111 AA;

Query Match 42.0%; Score 271.5; DB 22; Length 111;
Best Local Similarity 50.0%; Pred. No. 6.5e-18;
Matches 59; Conservative 13; Mismatches 35; Indels 11; Gaps 3;

QY 1 MSWGRSTYRPRRYVPEPMIGMRPEQFSDEVEPA--TPEGEPATQRQDPAAOE- 57

Db 1 MIWGRSTYRPRRSPVPPPELIGPW-----LEFGDEEPQOEPPFTESRDPAPQQR 52

QY 58 GEDEGASAGQGFKPEAHSQEOGHPTQGCEDGPDQCEMDPPNPPEVKTPEGEKQSQ 115

Db 53 EEDQGAARTQVDLEADLQELSQSKTGCGNGGPDQDKILPKSEQFKMPEGGDRQFQ 110

Search completed: April 7, 2003, 18:15:04

Job time : 36 secs

